UNIVERSIDAD DE GRANADA

Seminario 1

Herramientas bioinformáticas en Genómica

Biología Molecular - Grado en Medicina
Curso 2023/24

¿Qué vamos a hacer?

Vamos a usar bases de datos y herramientas bioinformáticas para conocer mejor un gen cuyas mutaciones causan una enfermedad.

Durante el seminario buscaremos información sobre el gen *KRAS*. Iremos respondiendo a las preguntas que he subido a PRADO y las entregaréis al final de la clase.

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

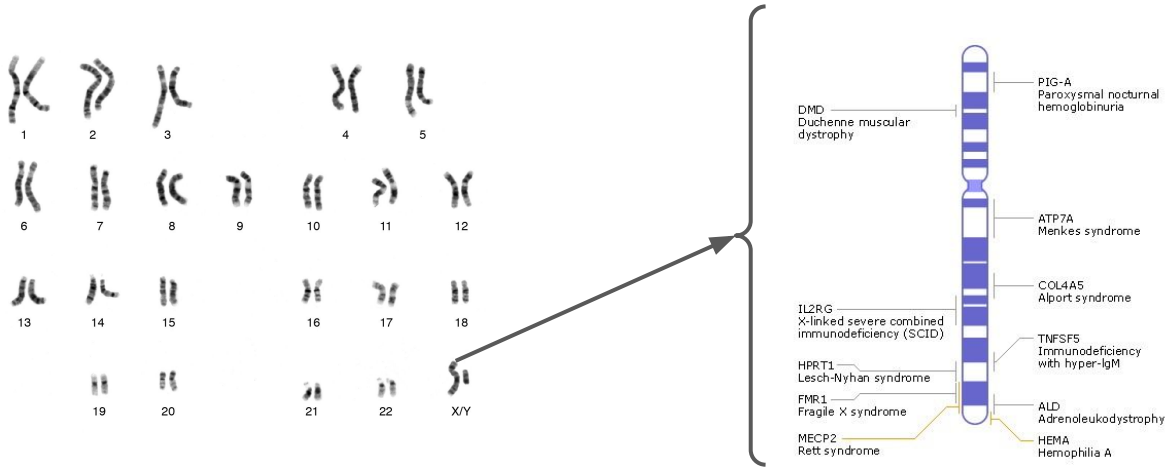
CUESTIONES SEMINARIO 1:
HERRAMIENTAS BIOINFORMÁTICAS EN GENÓMICA
Grado en Medicina | Biología Molecular | Curso 2023/24

Nombre y Apellidos:

1. Sobre el gen *KRAS*:
   - **a)** ¿Cuál es su Ensembl ID?
   - **b)** ¿En qué cromosoma, coordenadas y hebra se localiza?
   - **c)** Nombra sus 2 genes más próximos.
   - **d)** Indica el dbSNP ID y la consecuencia de una variante patogénica de este gen.
2. Sobre la isoforma (transcrito) de referencia del gen *KRAS*:
   - **a)** ¿Cuál es su Ensembl ID?
   - **b)** ¿Cuál es su longitud en pares de bases y en aminoácidos?
   - **c)** ¿Cuántos exones tiene?
3. Indica el identificador y la secuencia de un miRNA maduro que regule la expresión del gen *KRAS*.
4. ¿Qué laboratorio en España realiza el diagnóstico genético de mutaciones en KRAS?
5. Nombra otra línea celular distinta a NCI-H358 que contenga la mutación G12C en KRAS.
6. Indica la pareja de primers que has diseñado para amplificar el exón 2 completo de KRAS.

¿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.

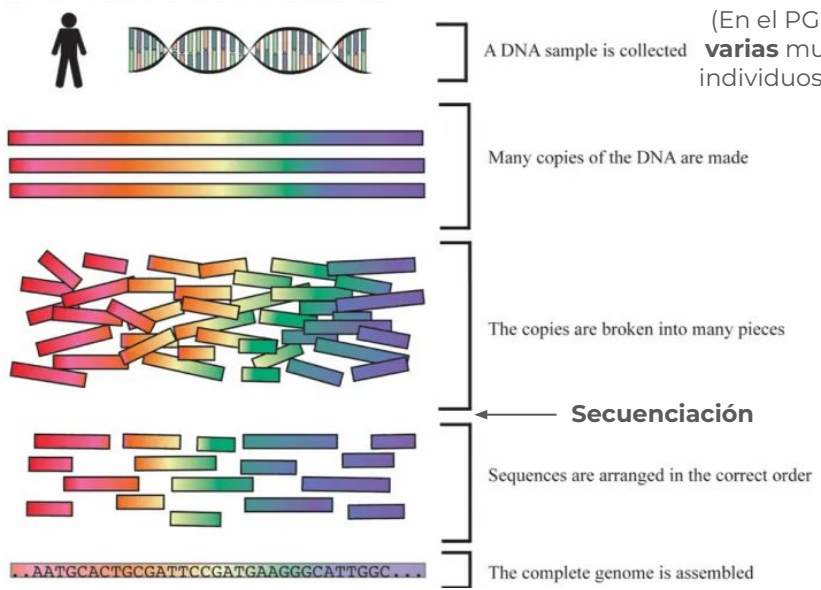


Proyecto Genoma Humano (1999-2003)

Secuenciar el genoma humano (3200 millones de pares de bases) y **publicar** los datos:

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

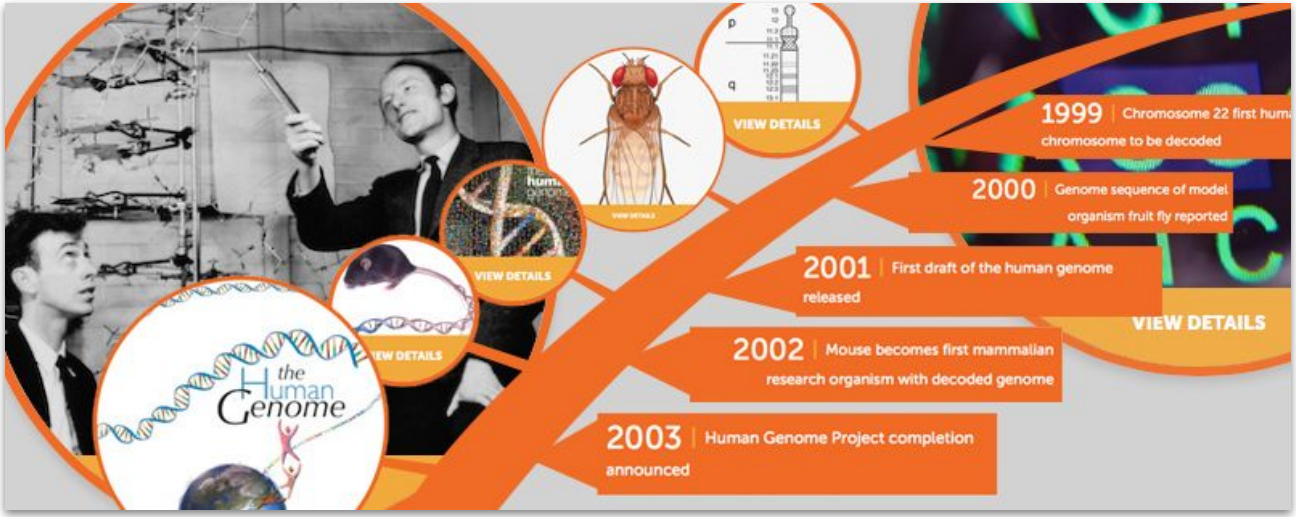
Proceso de secuenciación de un genoma



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

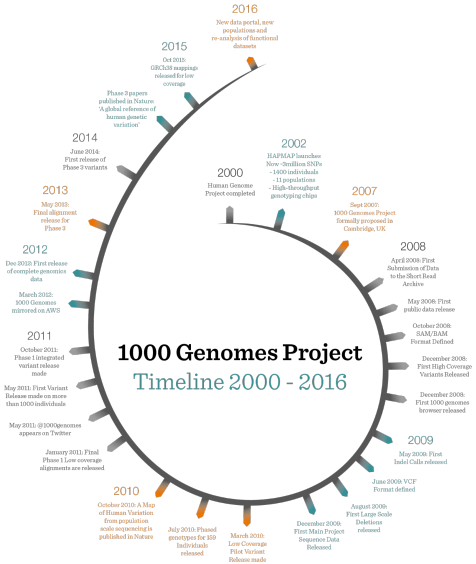
Proyecto Genoma Humano (1999-2003)

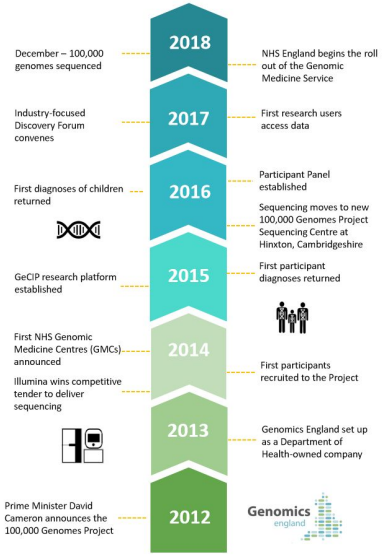
- El ADN secuenciado procedía de **varios individuos** donantes anónimos.
- La secuenciación se llevó a cabo entre **20 laboratorios colaboradores** de Estados Unidos, Reino Unido, Francia, Alemania, Japón y China y se completó en **4 años**.

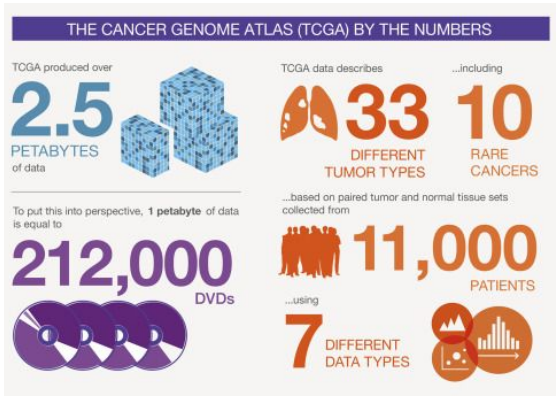


Nuevos proyectos genómicos

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







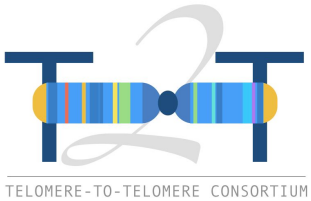
Proyecto **1000** genomas

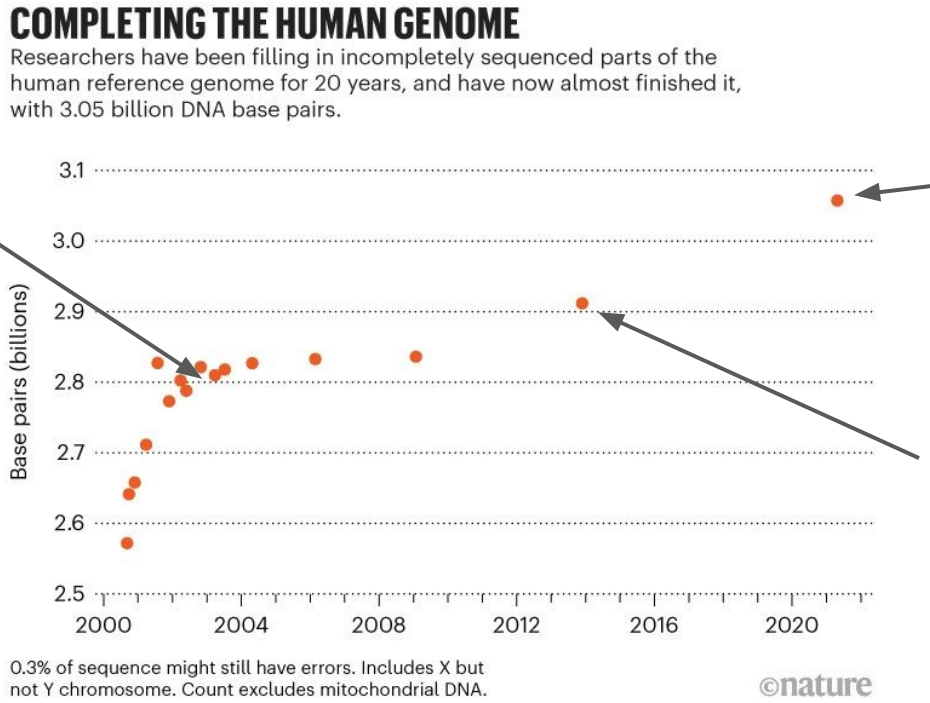
Proyecto **100000** genomas

Proyecto **The Cancer Genome Atlas** (TCGA)

Proyecto Telomere-to-Telomere (T2T, 2021)

Genoma secuenciado por el Proyecto Genoma Humano
(Año 2003, ~2800 millones de pares de bases)





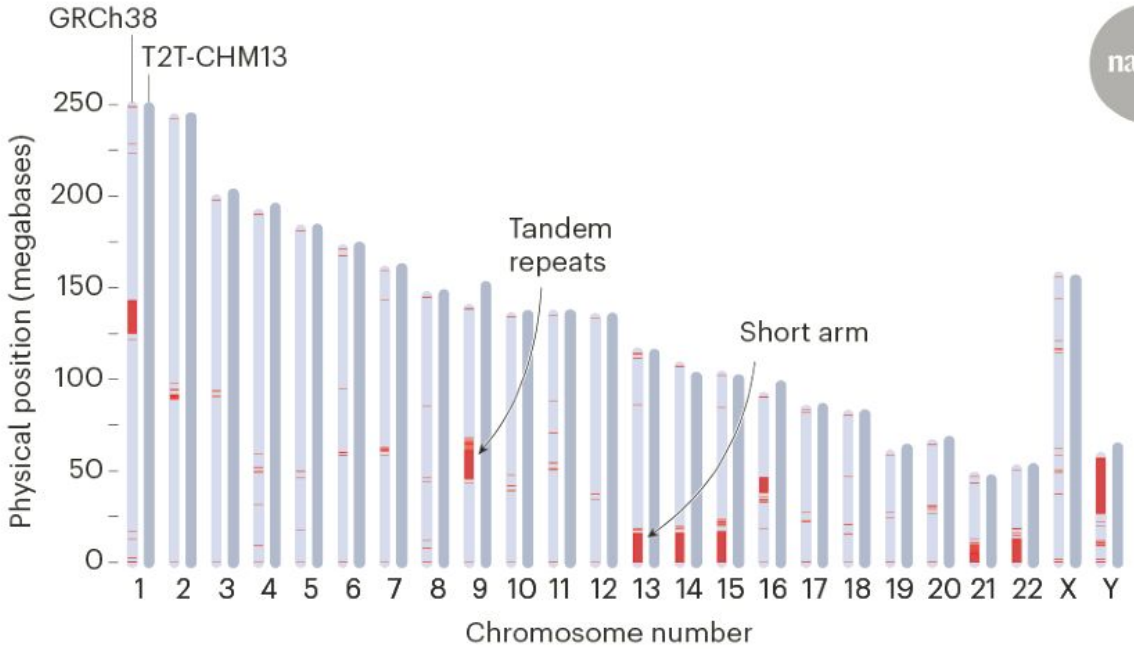
T2T-CHM13: Versión más completa del genoma humano
(Año 2021, ~3050 millones de pares de bases)

GRCh38: Última versión del genoma humano antes del proyecto T2T
(Año 2013, ~2900 millones de pares de bases)

Proyecto Telomere-to-Telomere (T2T, 2021)

T2T ha revelado regiones **repetitivas** del genoma hasta ahora desconocidas localizadas cerca de **centrómeros** y **telómeros**, principalmente.

TELOMERE-TO-TELOMERE CONSORTIUM



nature

Bases de datos

Nucleic Acids Research

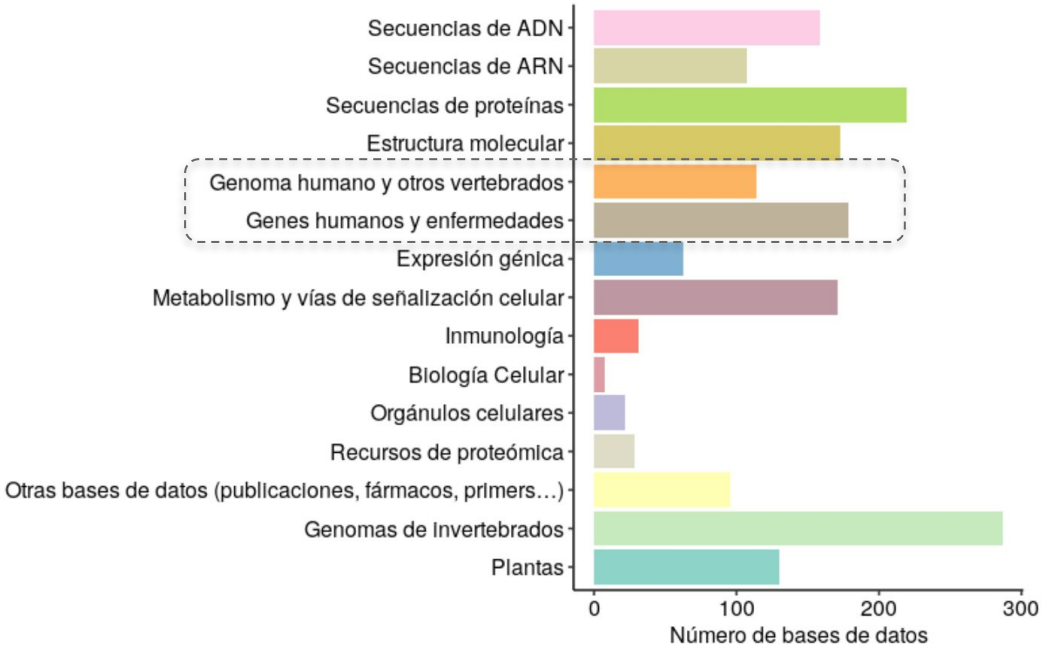
Nucleic Acids Research, 2024, **52**, D1–D9
https://doi.org/10.1093/nar/gkad1173
Advance access publication date: 30 November 2023
Editorial

The 2024 *Nucleic Acids Research* database issue and the online molecular biology database collection

Daniel J. Rigden[1,*] and Xosé M. Fernández[2]

- La revista *Nucleic Acids Research* recopila anualmente todas las bases de datos de Biología Molecular que existen (actualmente **1959**)
- En 2024 se han añadido **90** bases de datos nuevas.
- 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

mRNAWalk

UniProt

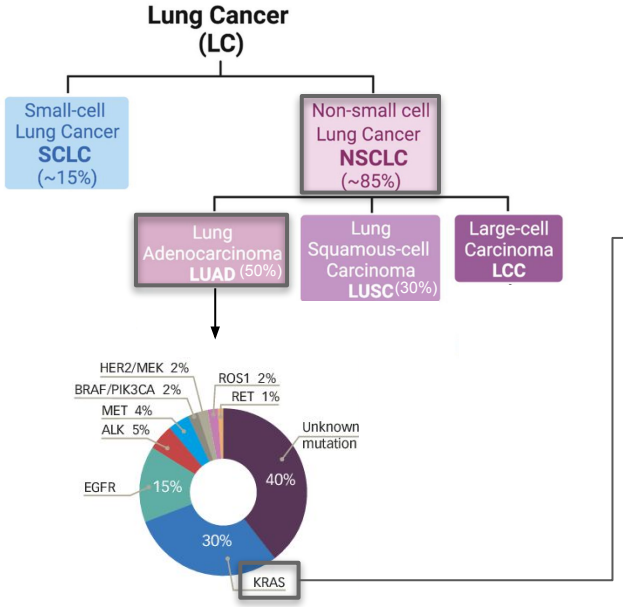
miRBase

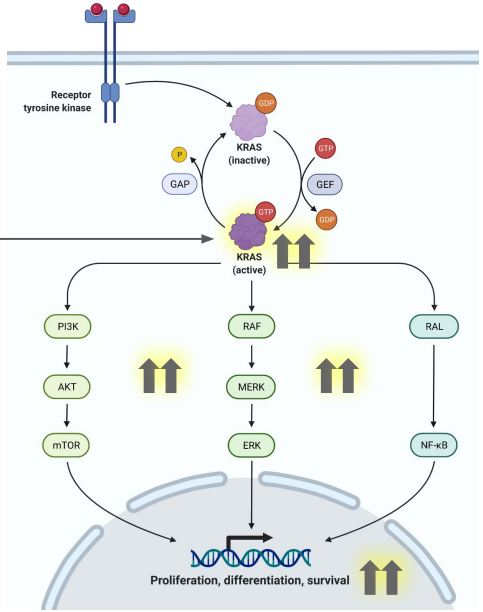
NCBI GTR Genetic Testing Registry

COSMIC Catalogue Of Somatic Mutations In Cancer

Gen *KRAS* en cáncer de pulmón

El **oncogen *KRAS*** está mutado en un **30%** de los casos de **adenocarcinoma** de pulmón (el tipo de cáncer de pulmón más frecuente).





Las mutaciones en KRAS hacen que esta proteína esté **más activa**, lo cual causa un **aumento descontrolado de la proliferación** celular.

La mayoría de mutaciones están en el **exón 2**. Las más frecuentes son dos mutaciones del aminoácido 12 (**G12C** y **G12D**).

Antes de empezar

Consideraciones para seguir mejor la práctica:

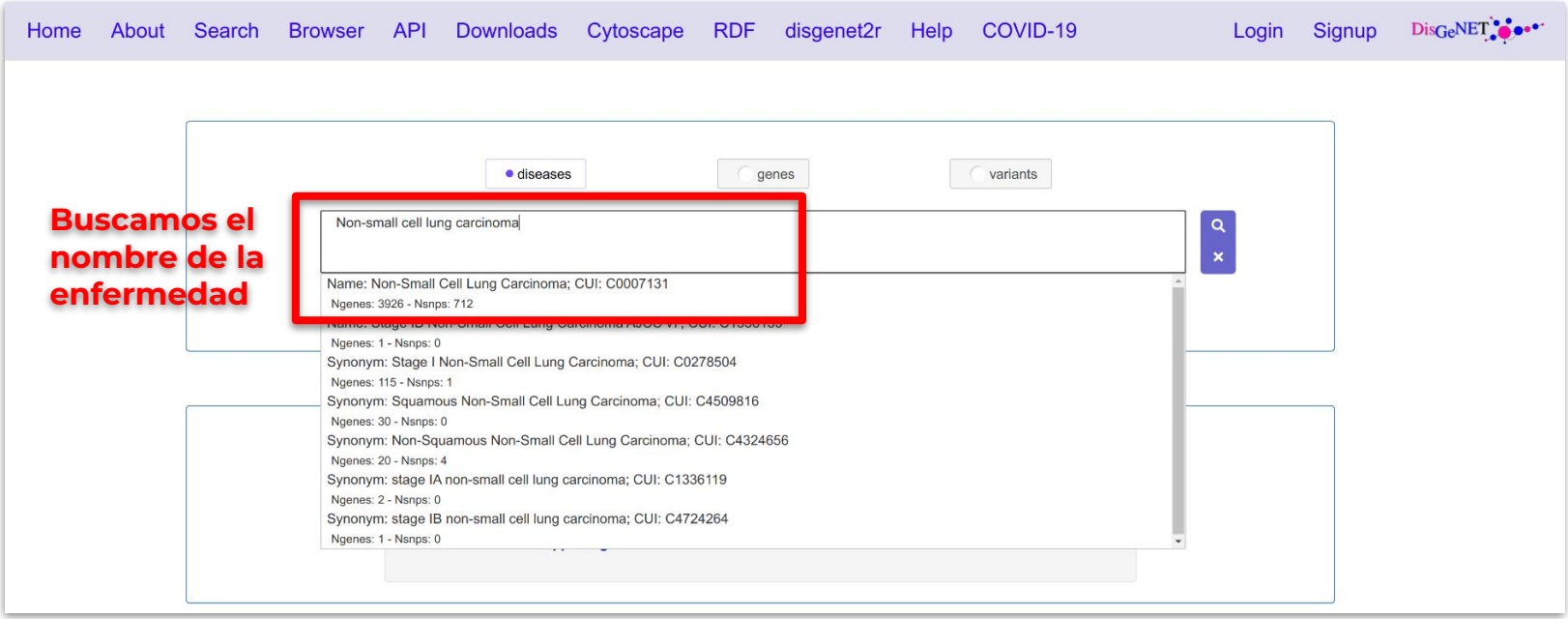
- Los **rectángulos rojos** de las diapositivas indican dónde hay que escribir o hacer click en las páginas web.
- La información oculta con **recuadros blancos** son las respuestas a las preguntas de la entrega.
- Si en algún momento te pierdes (¡es totalmente normal!), levanta la mano.

¿Hay algún gen asociado a esta enfermedad?

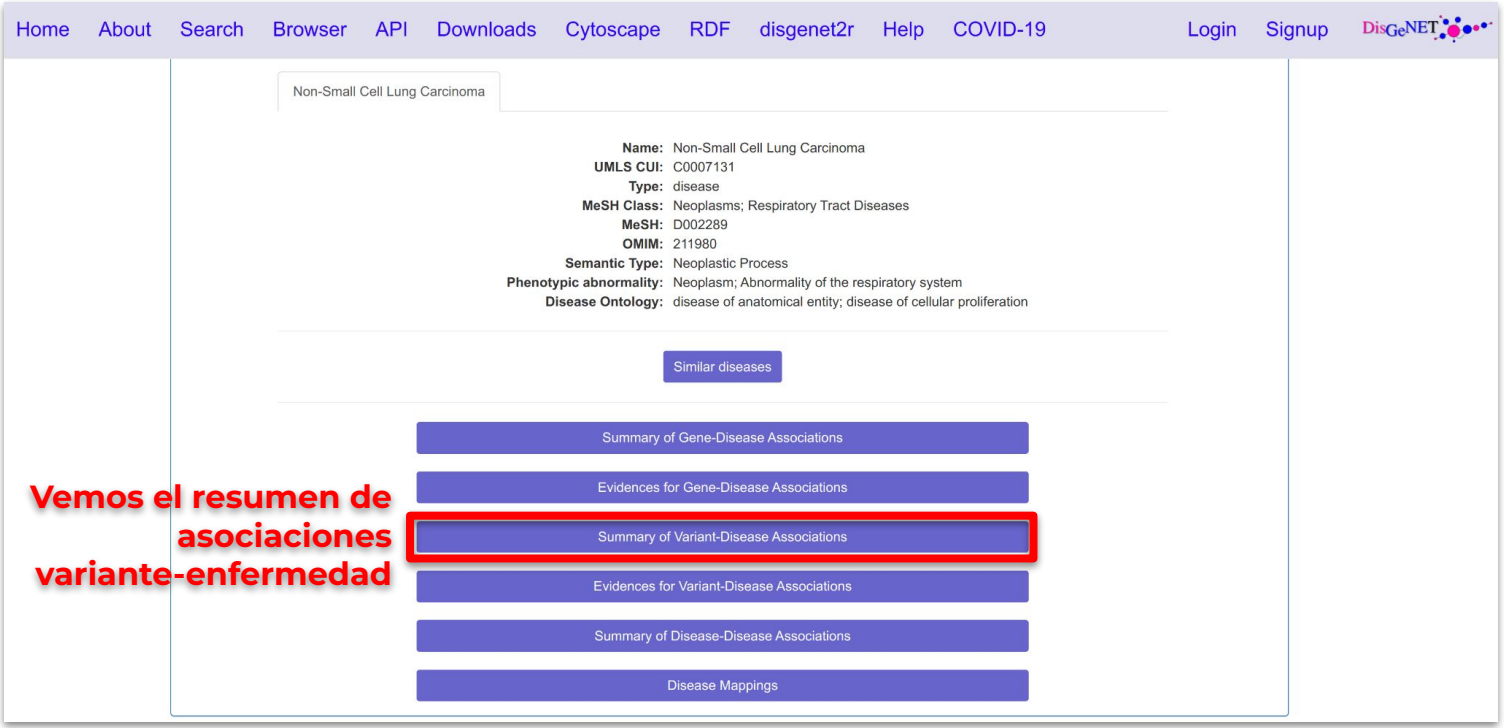
Respuesta: 

Enlace

Buscar en DisGeNet

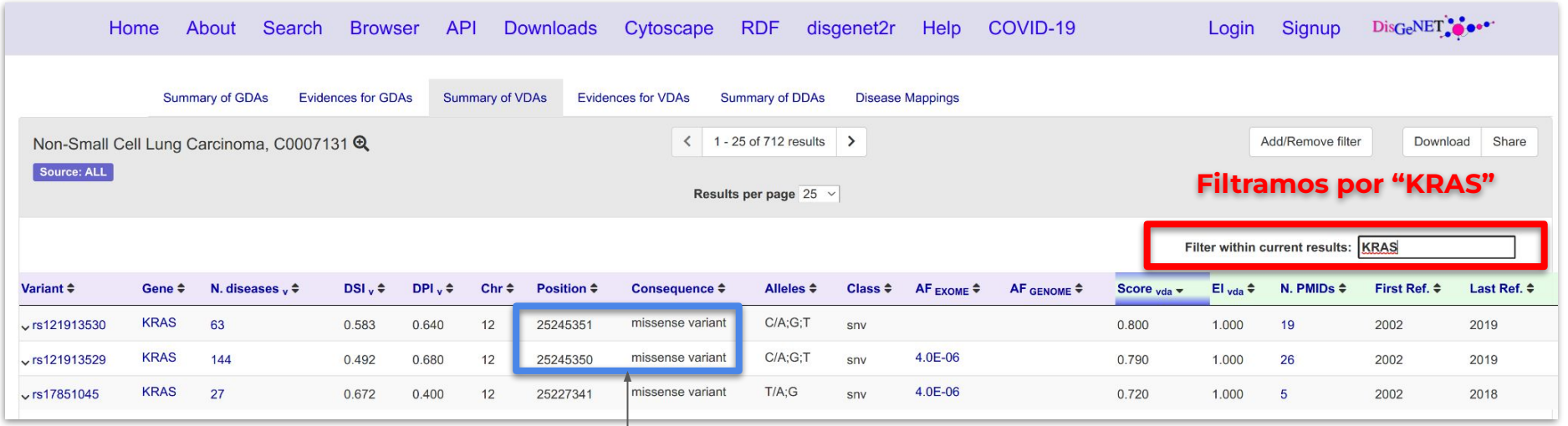


Asociaciones mutación-enfermedad



Vemos el resumen de asociaciones variante-enfermedad

Asociaciones mutación-enfermedad



Las dos mutaciones de *KRAS* más frecuentes (**G12C** y **G12D**)

¿Qué características tiene el gen?

Respuesta: Ensembl

Enlace

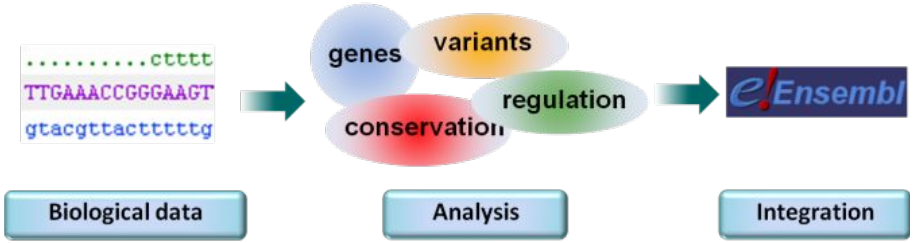
Preguntas 1 y 2

Ensembl

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

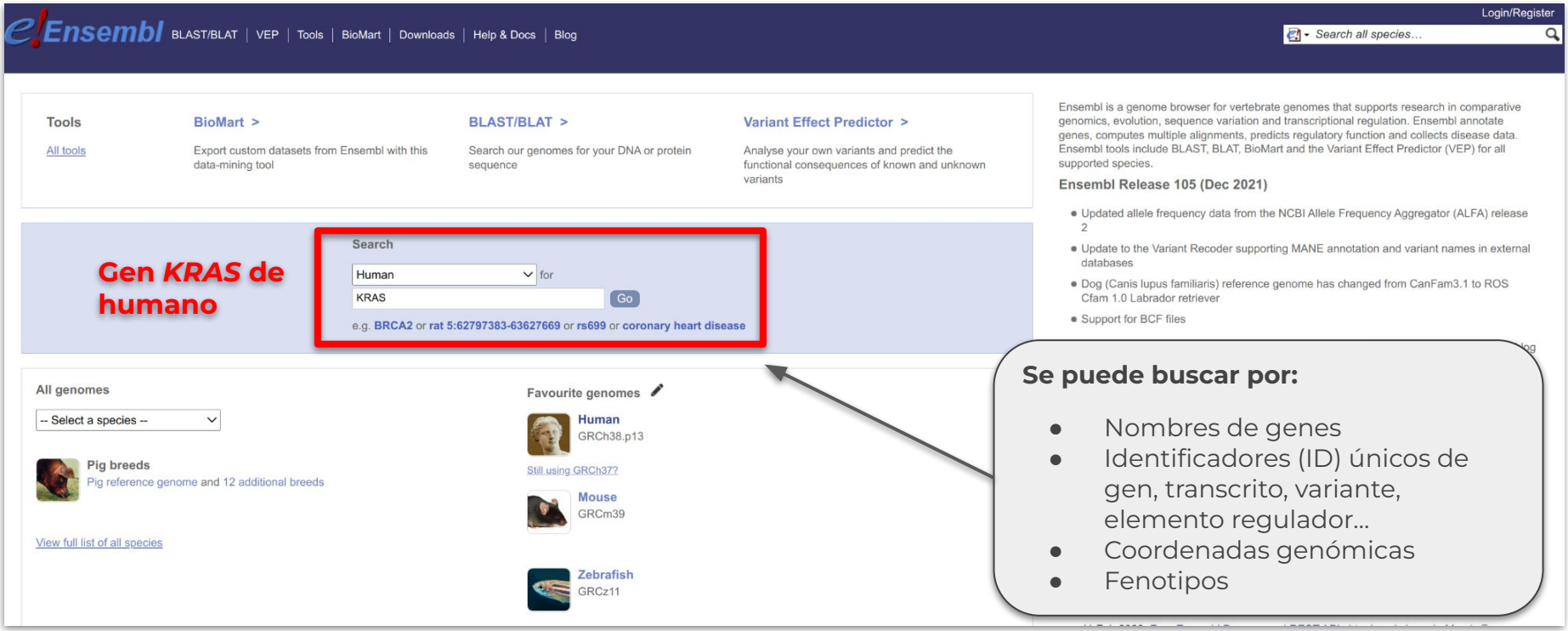
Contiene información de genes y transcritos a nivel de genoma, gen y proteína:

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

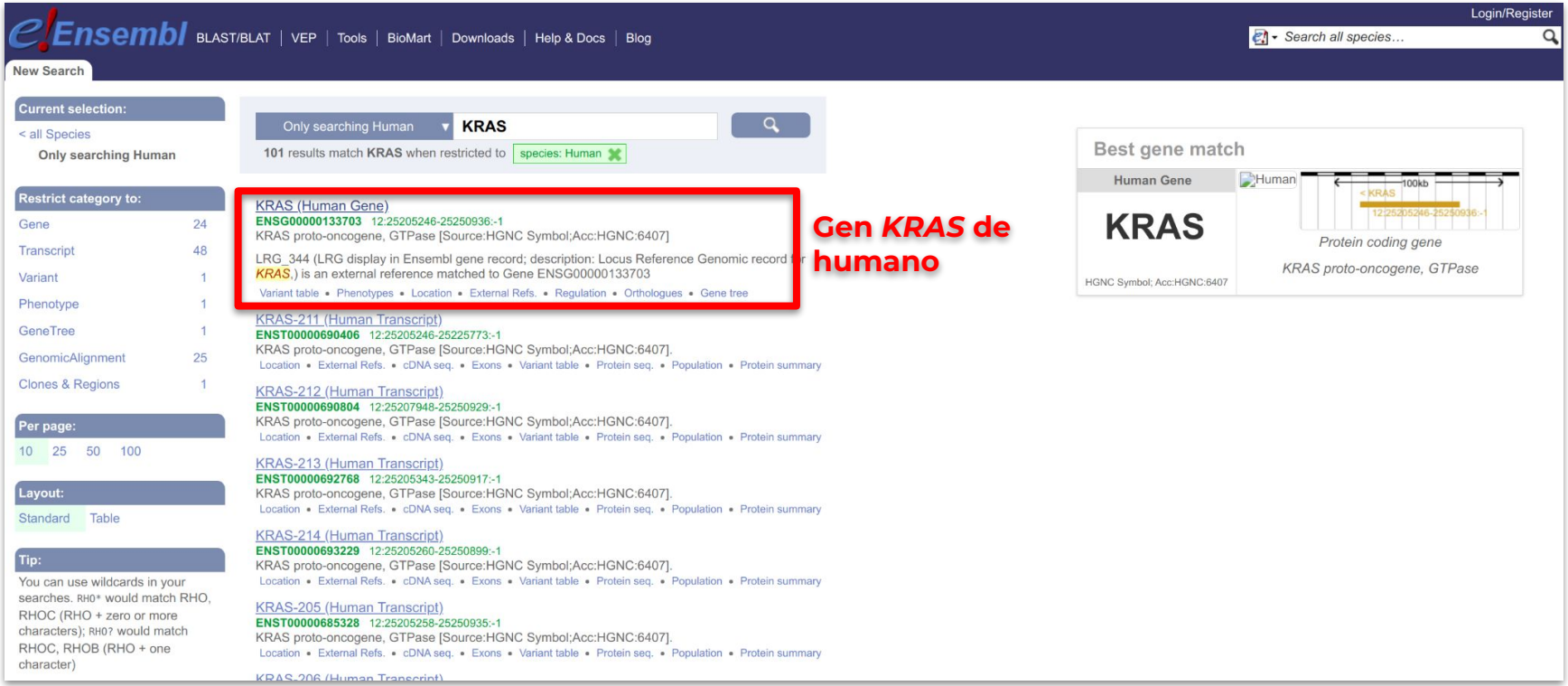


Otro navegador genómico interesante es UCSC Genome Browser.

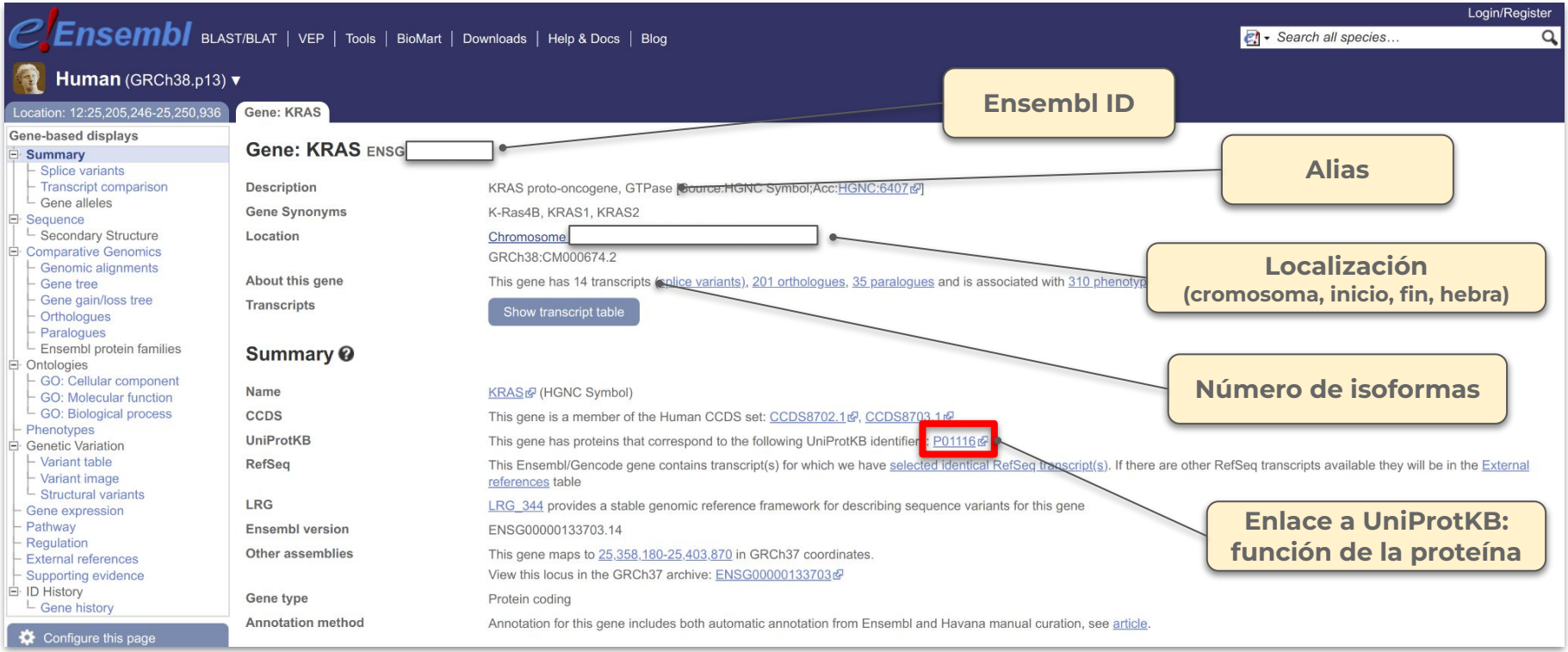
Buscar en Ensembl



Buscar en Ensembl

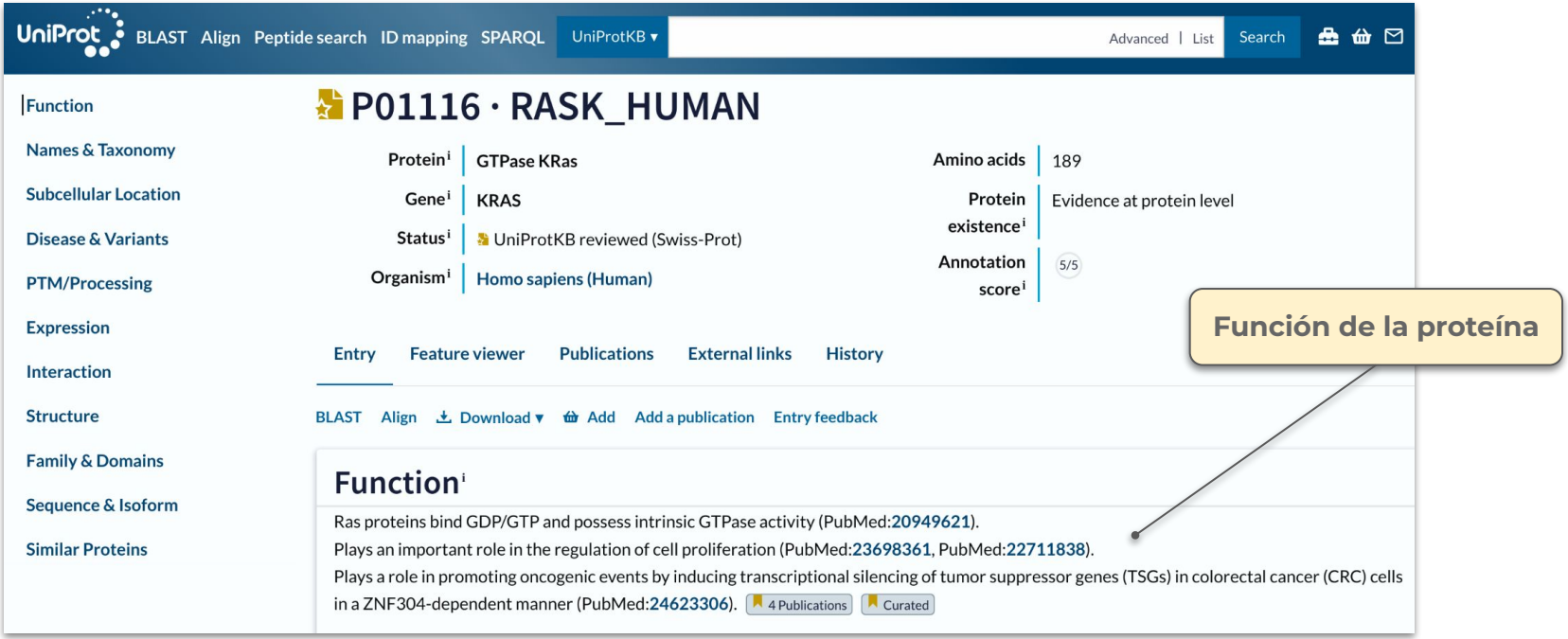


Características del gen



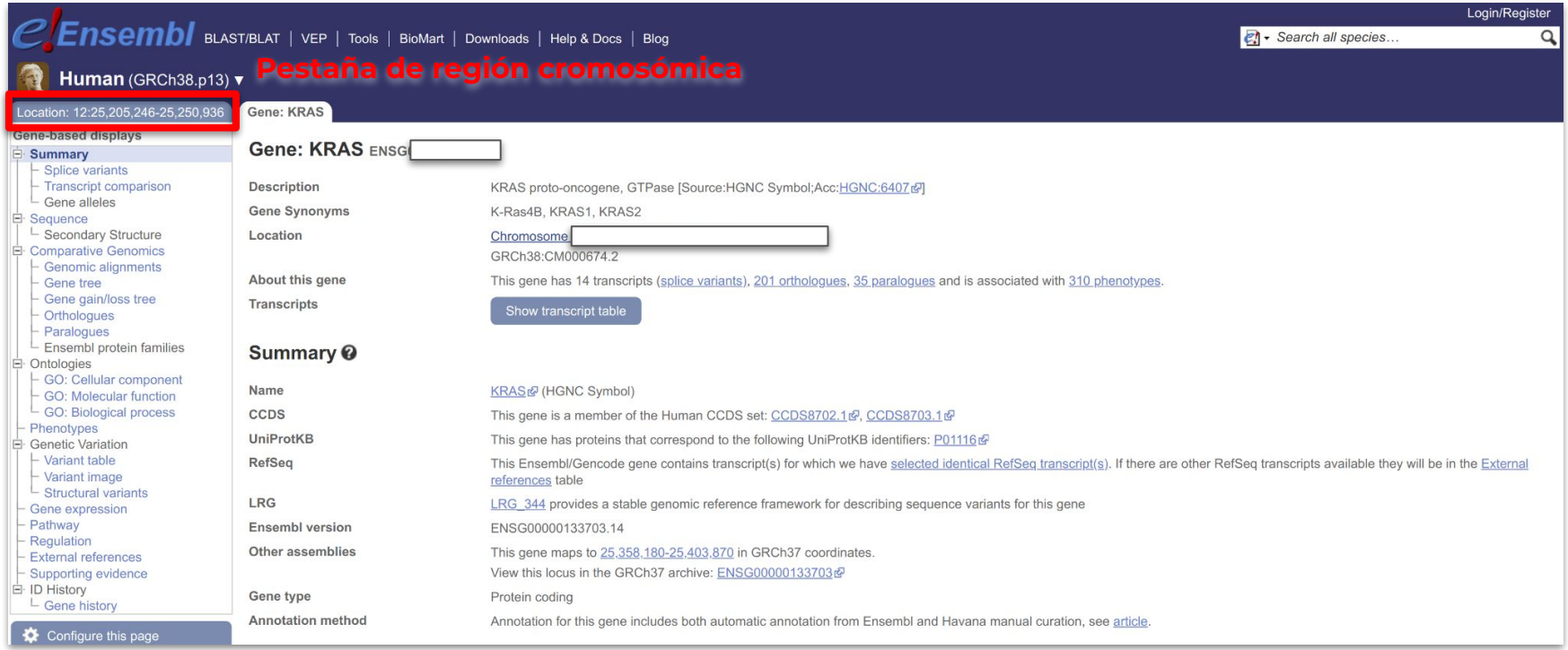
Función del gen (UniProtKB)

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

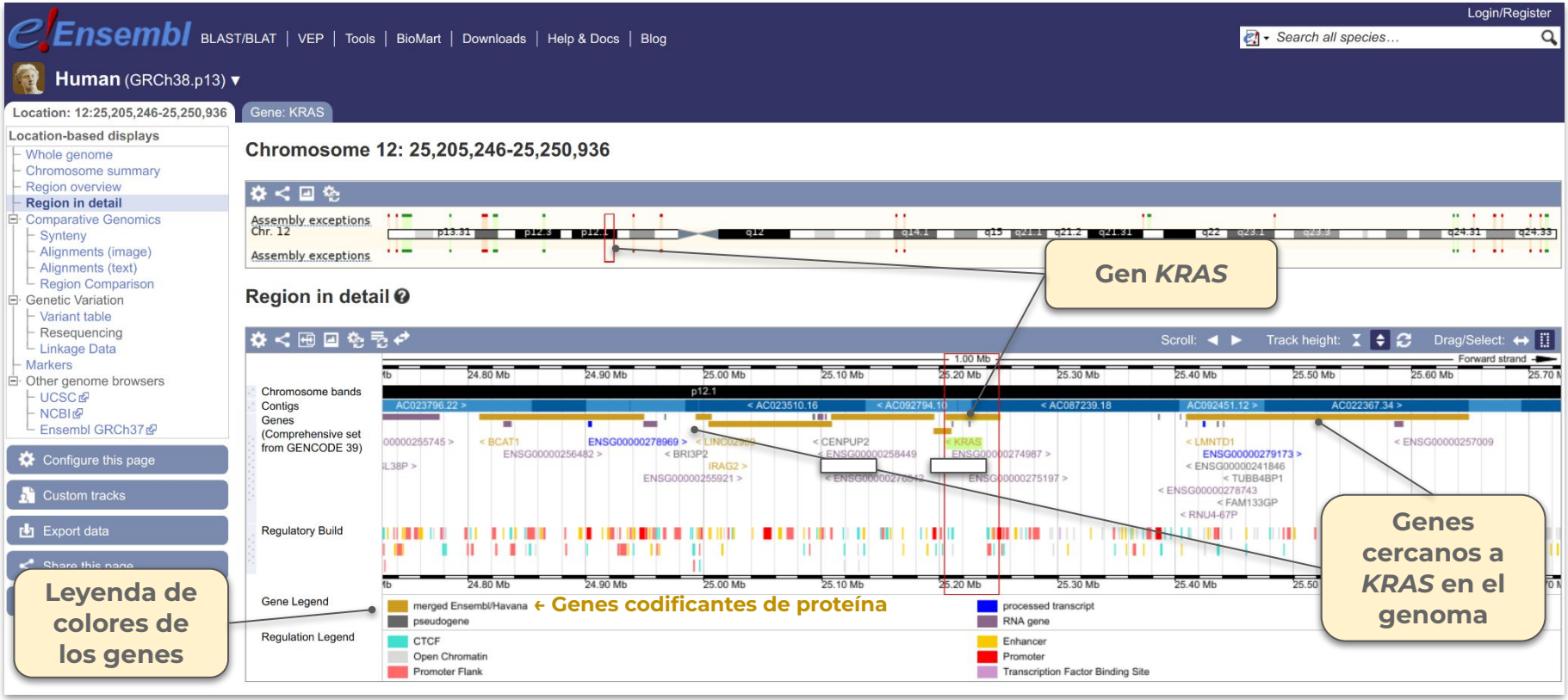


Genes vecinos

Volvemos a Ensembl:



Genes vecinos



Pistas de Ensembl

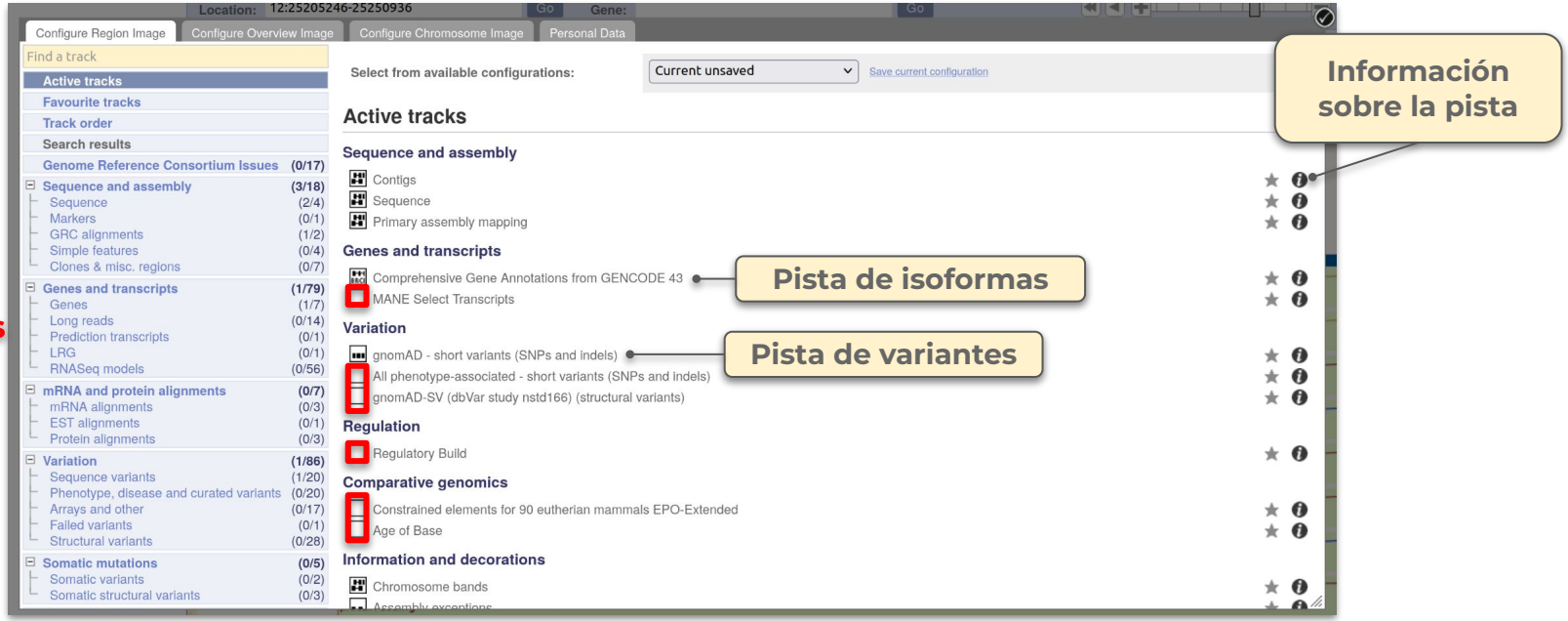
Hacemos scroll en la pantalla anterior hasta llegar a este otro visualizador genómico:

- 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:

- **Paso 1**: ocultar las pistas que no nos interesan.



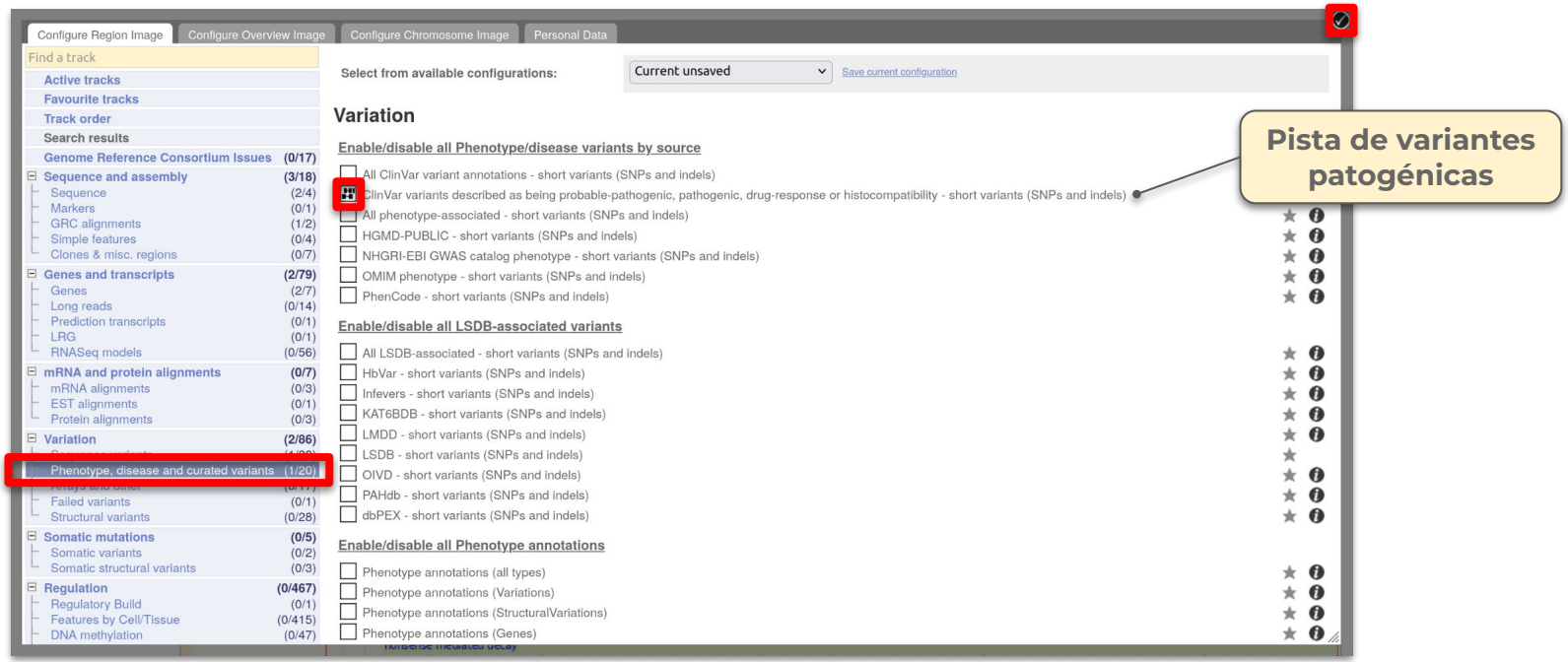
Variantes patogénicas

Queremos visualizar sólo las isoformas y las variantes:

- **Paso 2**: mostrar la pista de variantes patogénicas (estaba oculta).

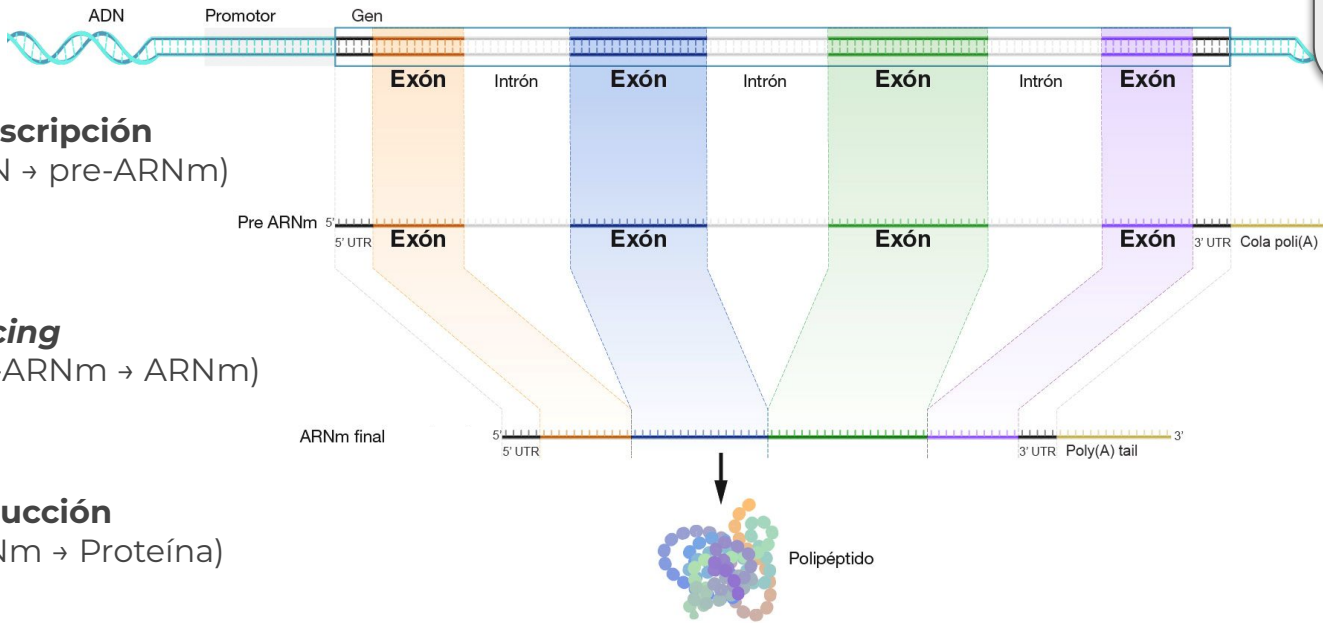
Activamos la pista de variantes patogénicas

(en modo Normal)



Recordatorio: estructura de los genes

Representación de un gen en **Ensembl**:



Leyenda:

- UTRs
- Exones codificantes
- Intrones

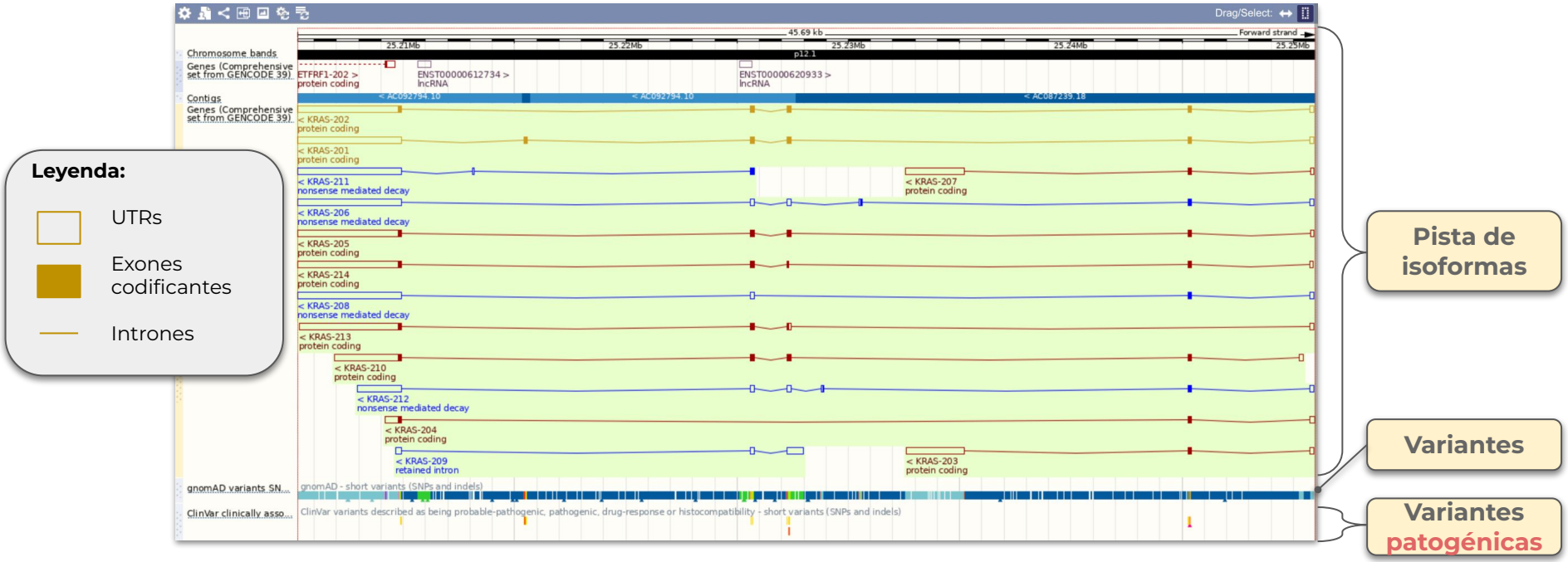
Transcripción
(ADN → pre-ARNm)

Splicing
(pre-ARNm → ARNm)

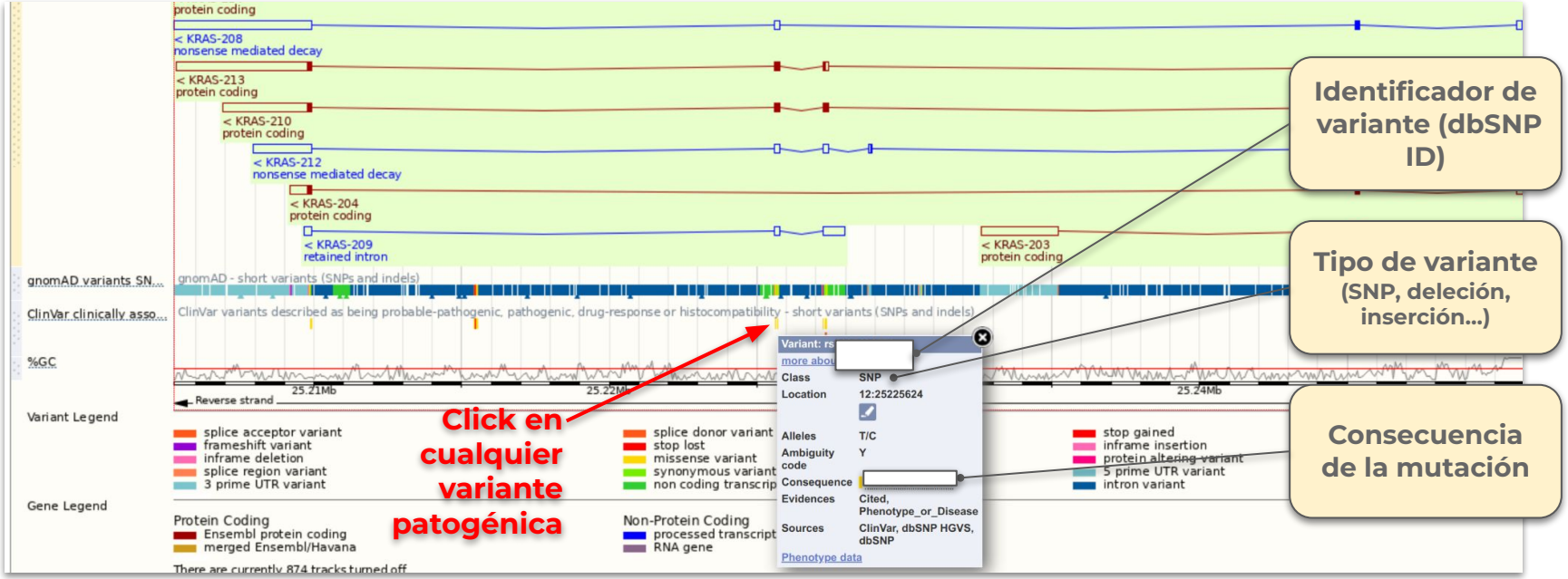
Traducción
(ARNm → Proteína)

Variantes patogénicas

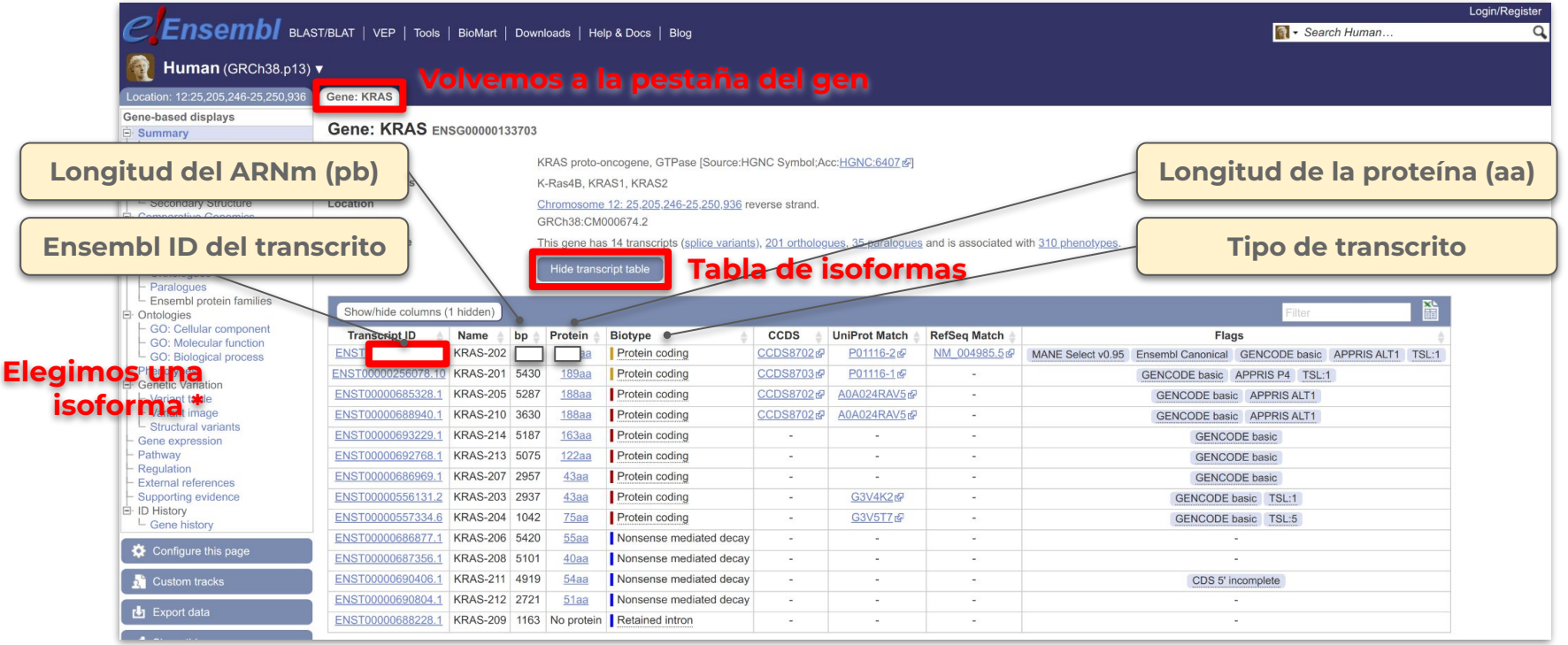
Hay variantes en todo el gen, pero las **patogénicas** están principalmente en **exones**.



Variantes patogénicas

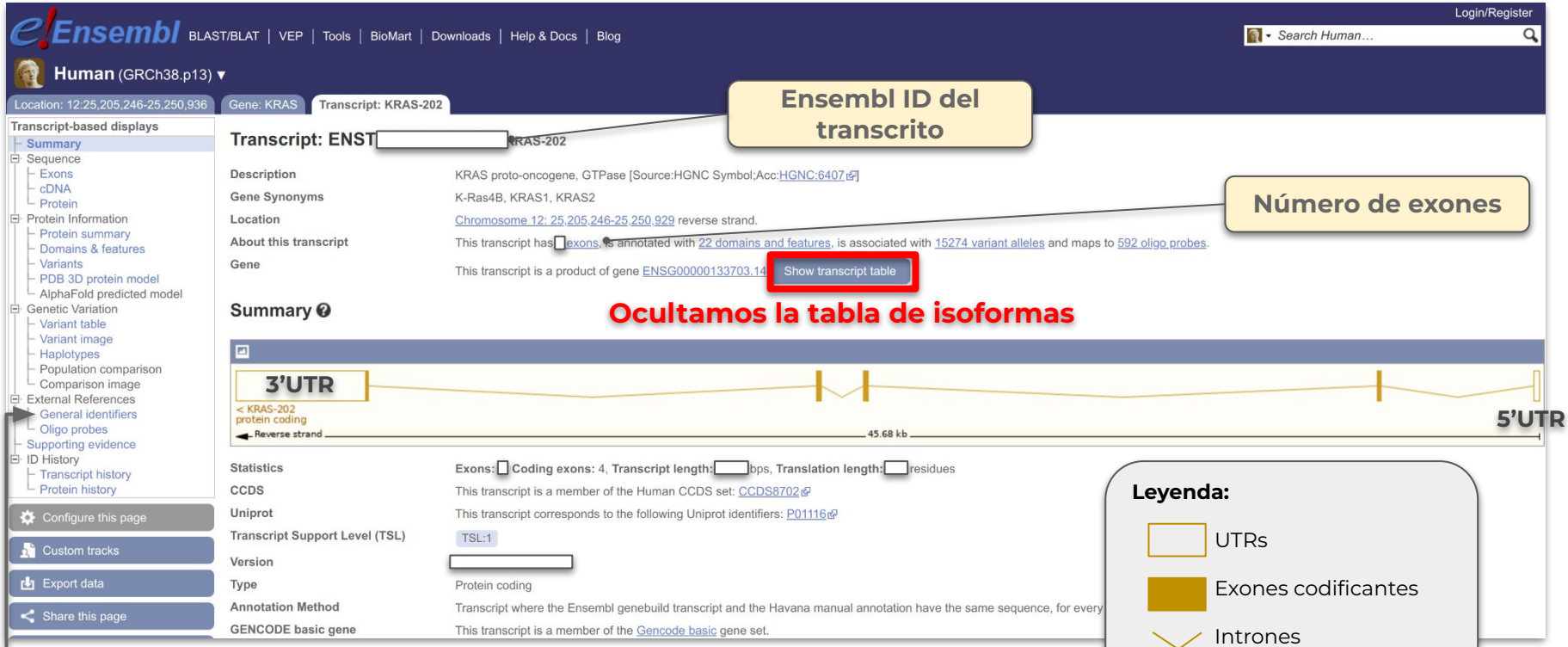


Isoformas



***** Preferiblemente la isoforma con la etiqueta "**MANE Select**" (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.

¿Existe algún microARN que regule al gen?

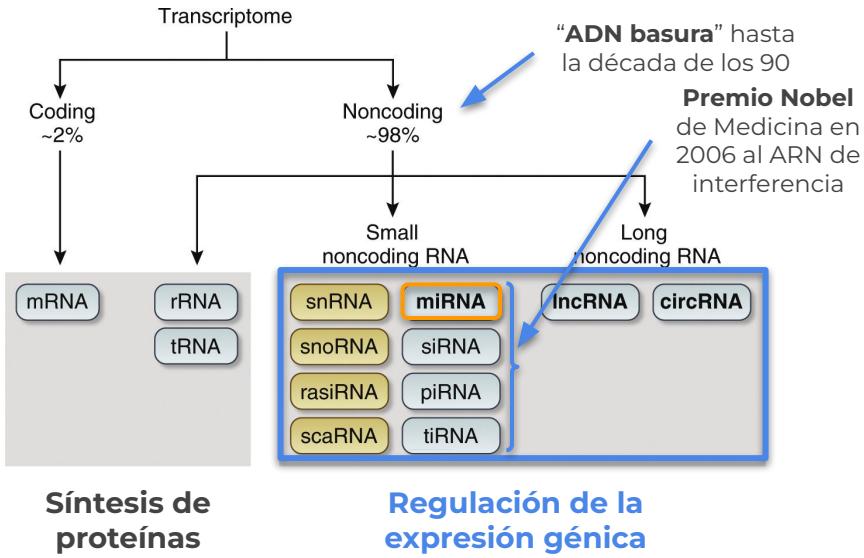
Respuesta: 

Enlace

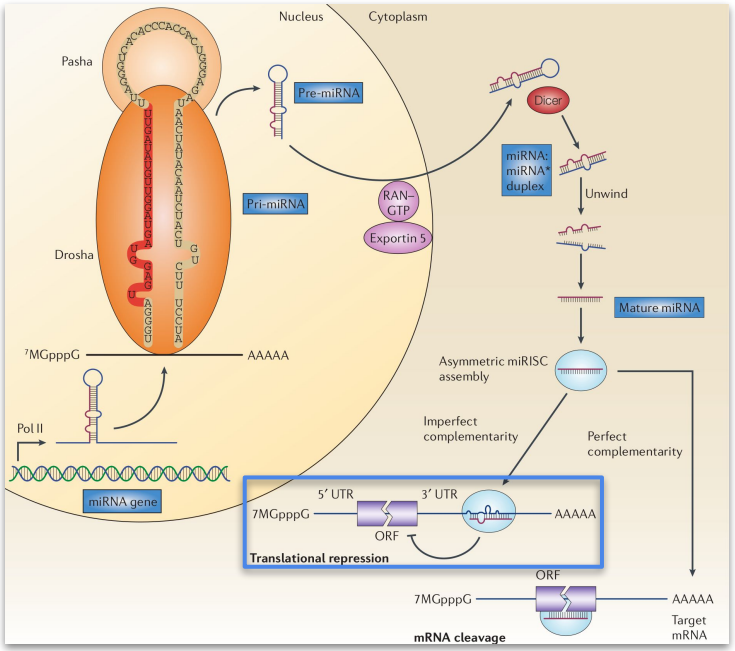
Pregunta 3

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

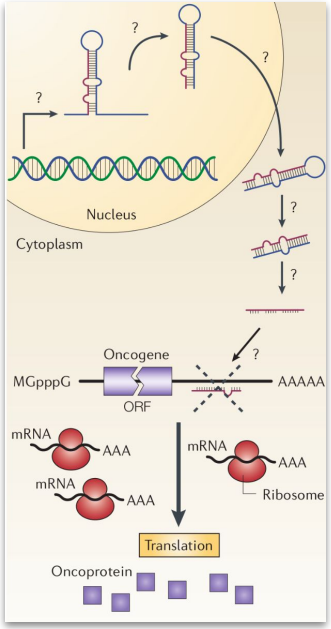


Biogénesis y actuación de miRNA

- 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)

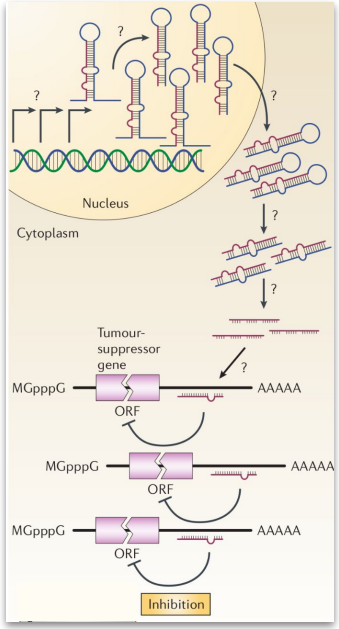


Si se infraexpresa

Si se sobreexpresa

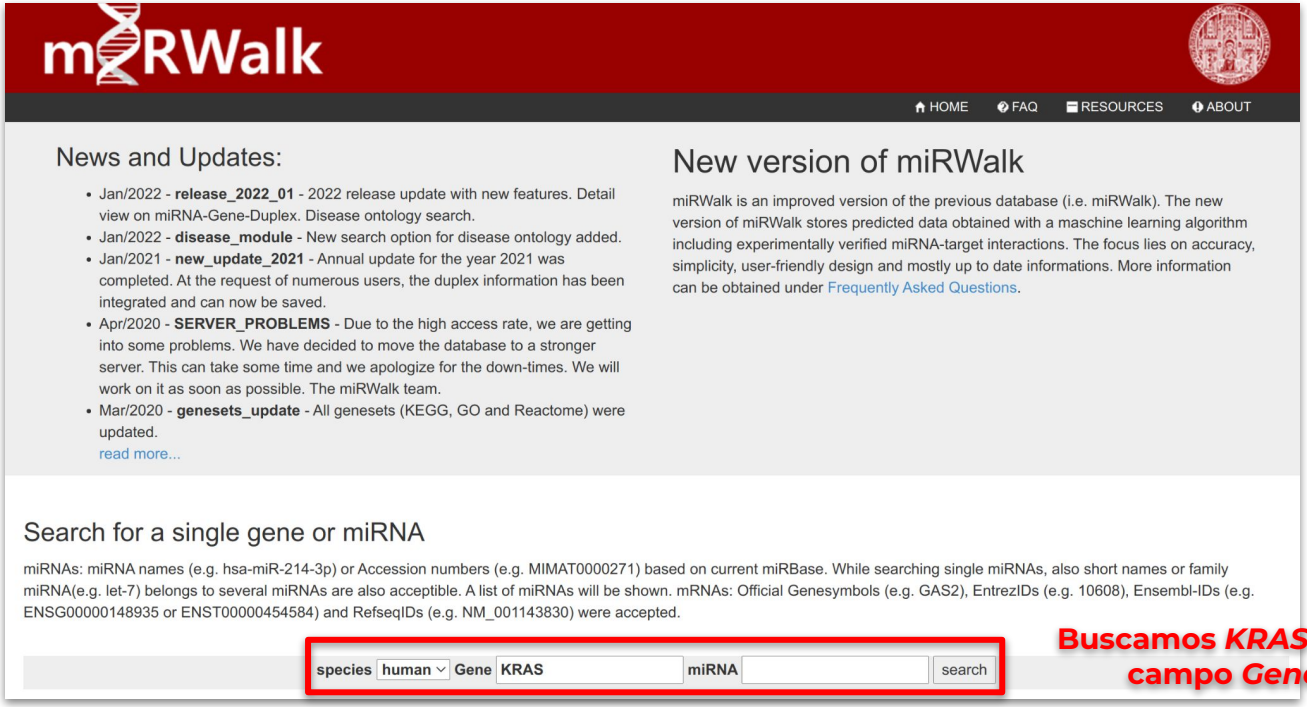
Formación de un tumor
↑ Proliferación
↑ Invasión
↑ Angiogénesis
↓ Muerte celular

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



miRWalk

Base de datos sobre relaciones entre miRNAs y genes.



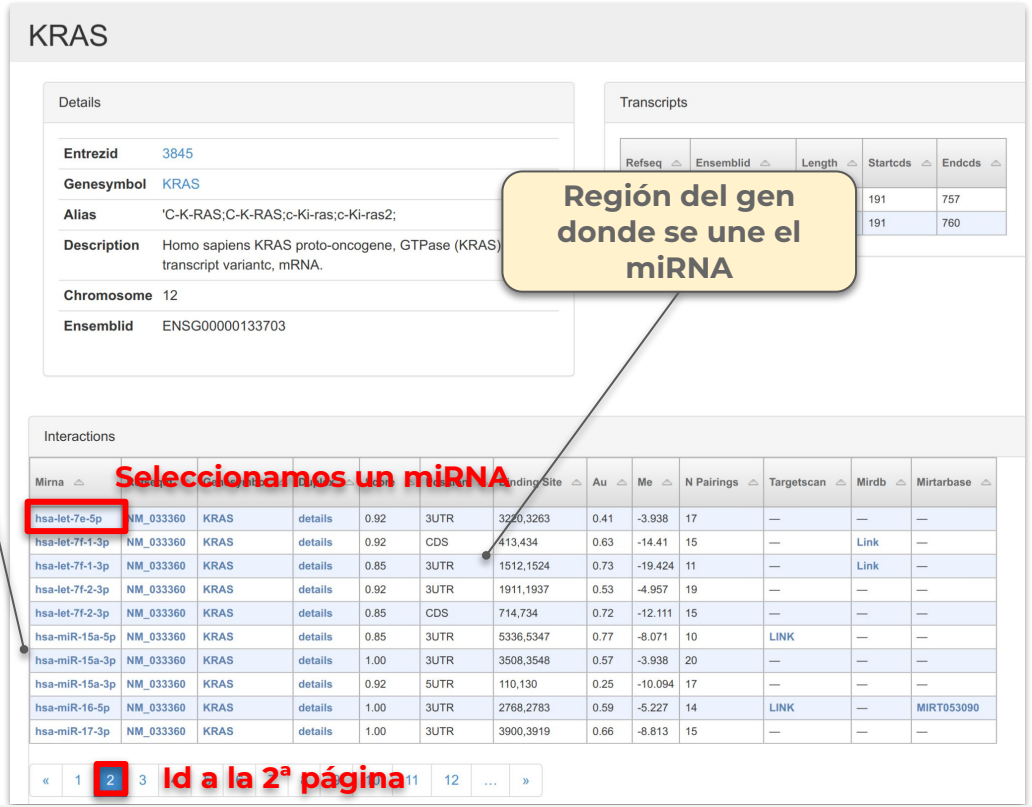
Buscamos *KRAS* en el campo *Gene*

miRNA asociados al gen

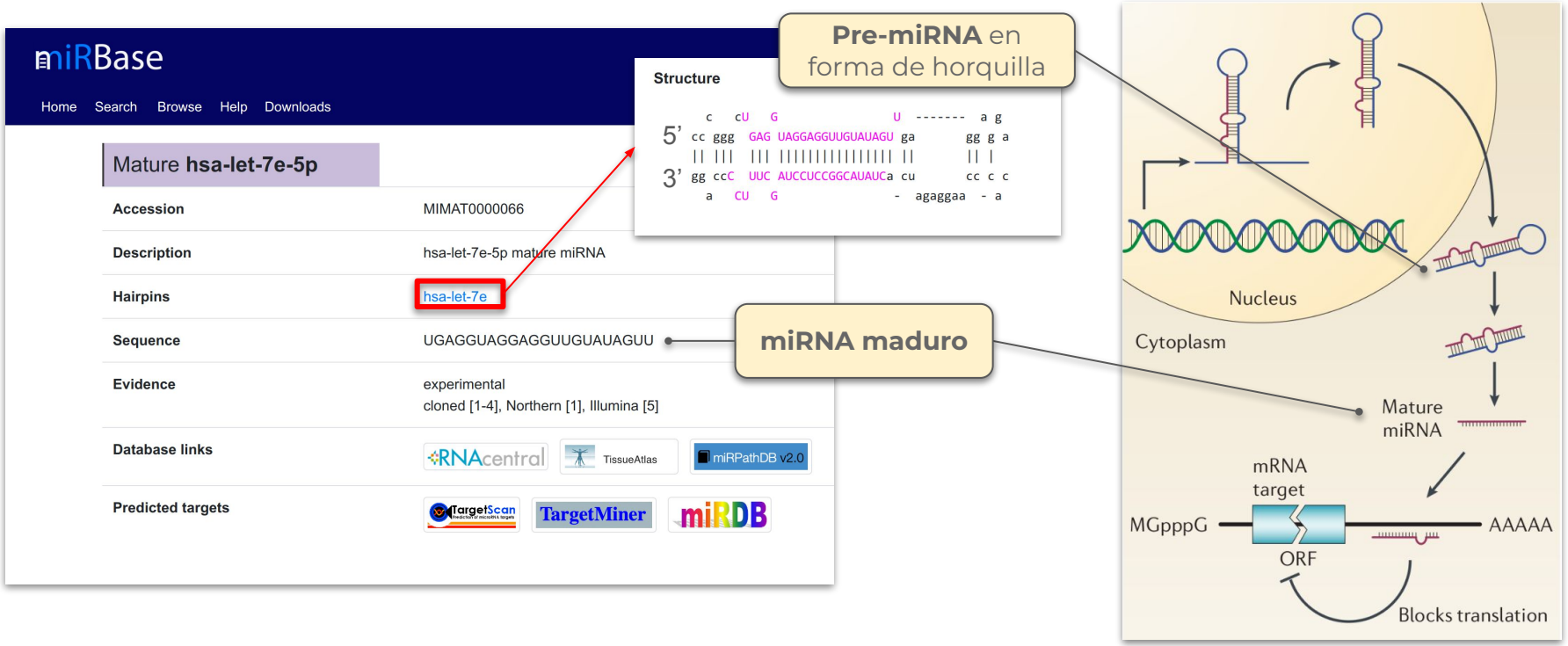
Identificador del miRNA

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)
- **let-7 es una excepción, se descubrió antes de que existiera esta nomenclatura**



Información del miRNA en miRBase



¿Qué laboratorios realizan el diagnóstico genético en España?

Respuesta: NCBI GTR Genetic Testing Registry

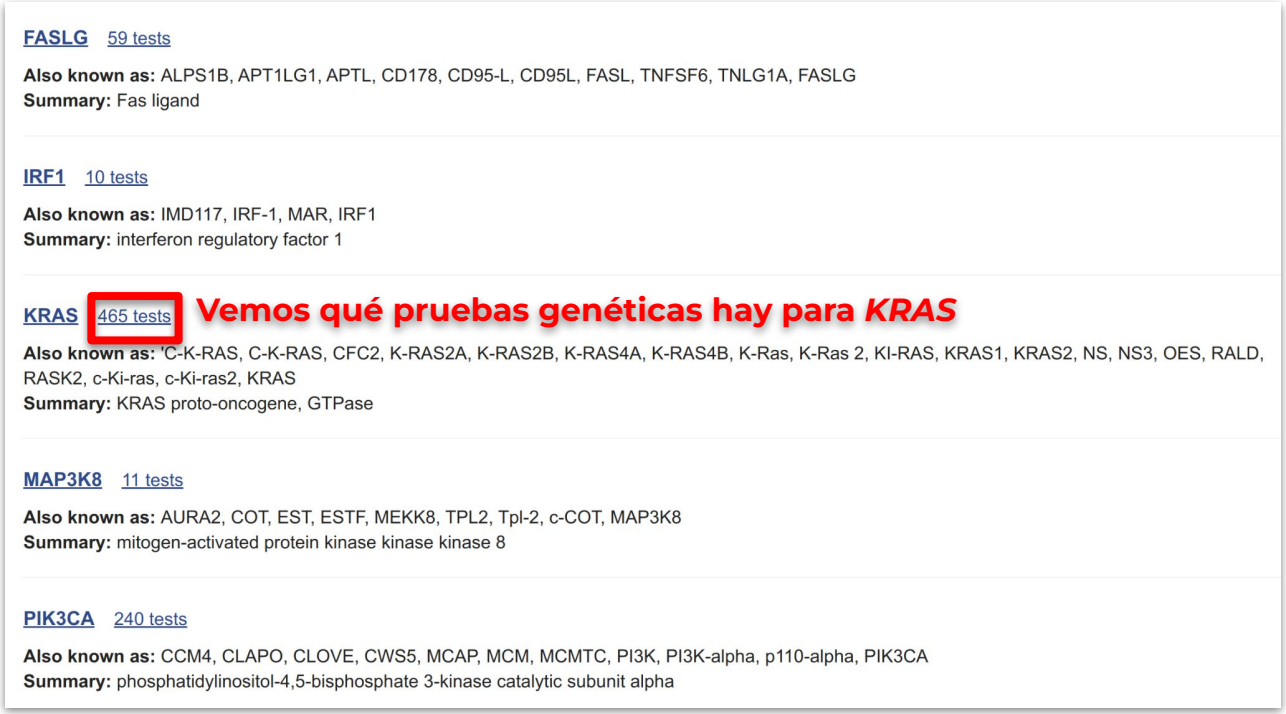
Enlace

Pregunta 4

Buscar en NCBI GTR

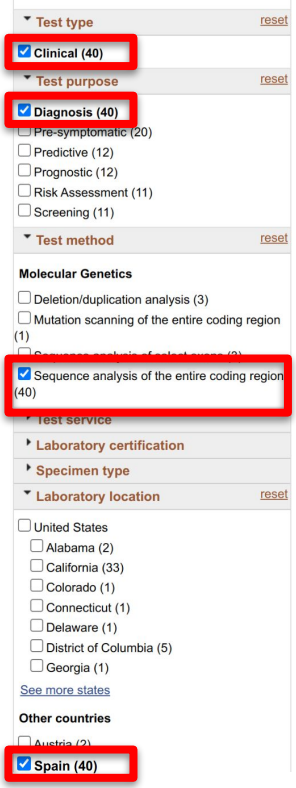
Buscar en NCBI GTR

Haced scroll al apartado de “Genes” hasta que encontréis el gen *KRAS*:



Tests genéticos de *KRAS* en España

Aplicamos 4 filtros



Genes analizados en la prueba

Nombre del laboratorio

Las pruebas que analizan varios genes se conocen como **paneles de secuenciación**.

Si sólo estamos interesados en secuenciar el gen *KRAS*, buscaríamos una prueba que analice sólo este gen.

Preparación para la Práctica 2: Amplificación de fragmentos de ADN por PCR

1. Elección de líneas celulares con *KRAS* mutado.
2. Diseño de *primers* para amplificar el exón 2 de *KRAS*.

¿Cómo encuentro una línea celular de cáncer de pulmón con *KRAS* mutado?

Respuesta: COSMIC Cell lines
Catalogue Of Somatic Mutations In Cancer

Enlace

Pregunta 5

¿Qué es una línea celular?

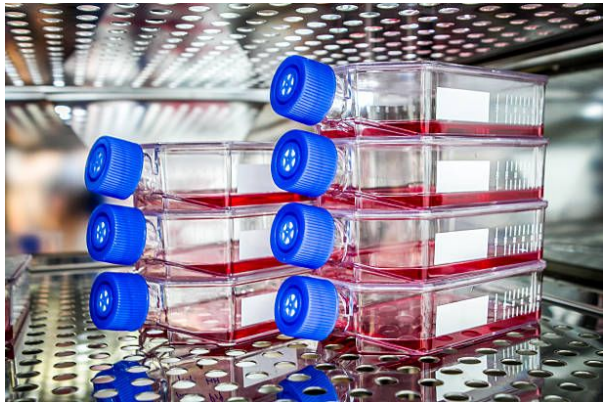
Es un cultivo de células establecido que **proliferará indefinidamente** si las condiciones del **medio** de cultivo y **espacio** son adecuadas.

- Se usan en **investigación** como paso previo a los ensayos en organismos modelo (ratones, ratas, etc)
- Hay líneas celulares de distintos tipos de cáncer y tejidos sanos.
- Una línea celular humana proviene de un donante:

  La línea celular **HeLa** de cáncer de cuello de útero proviene de una muestra del tumor de Henrietta Lacks, tomada en 1951.

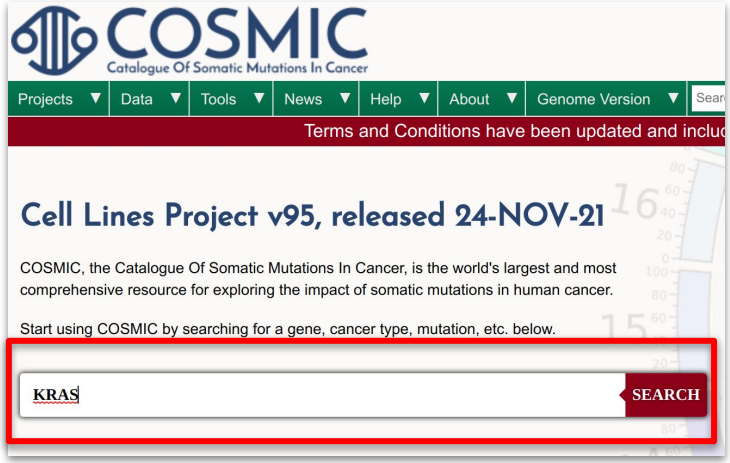
En la **Práctica 2** usaréis estas líneas de cáncer de pulmón:

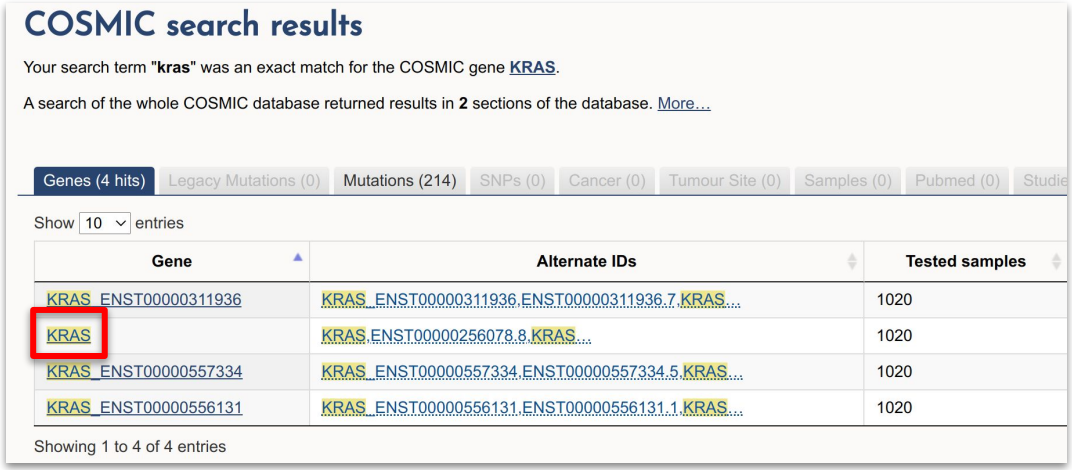
- *KRAS* mutado G12C: **NCI-H358**
- *KRAS* mutado G12D: **A427**
- *KRAS* no mutado: **NCI-H838**



Frascos de cultivos celulares en un incubador.

Buscar en COSMIC Cell Lines





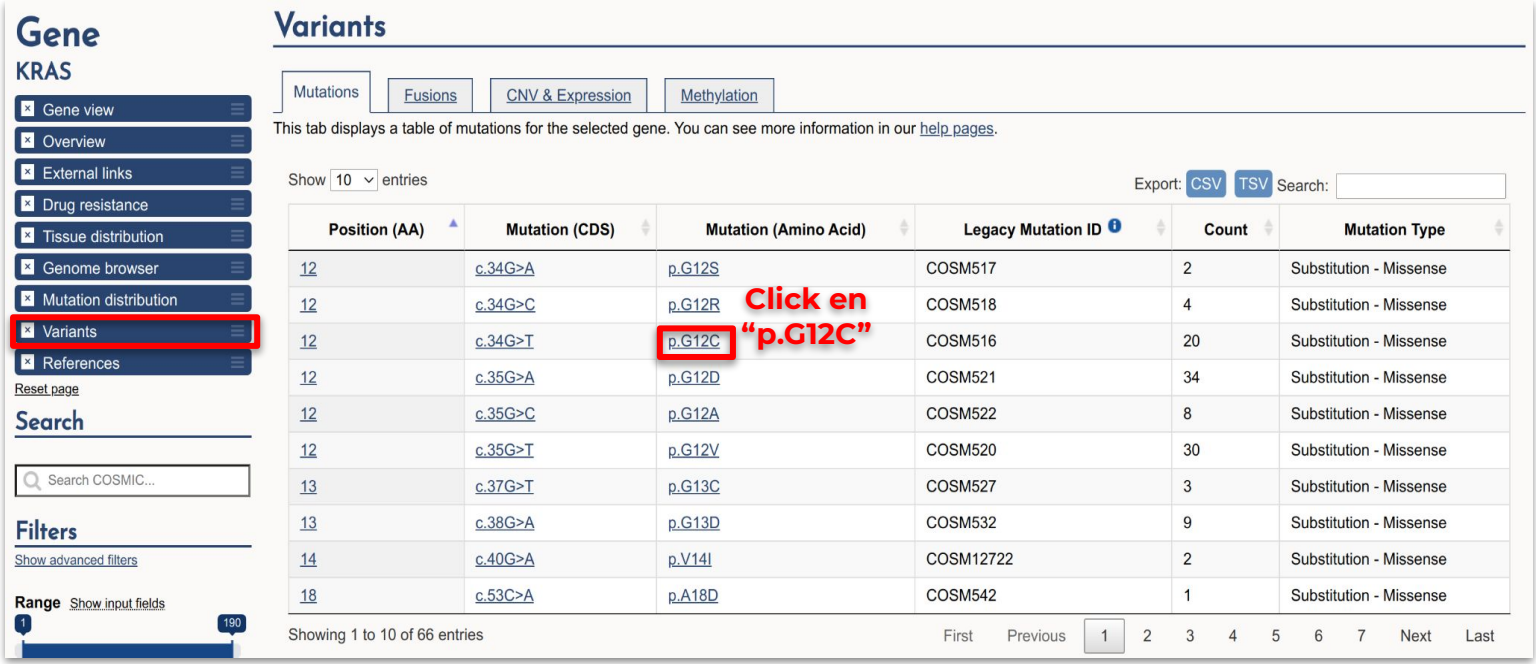
Escribimos *"KRAS"* en el buscador

Hacemos click en el segundo resultado

Buscar en COSMIC Cell Lines

Vamos a buscar líneas celulares con la mutación **G12C** en *KRAS*:

Click en "Variants"



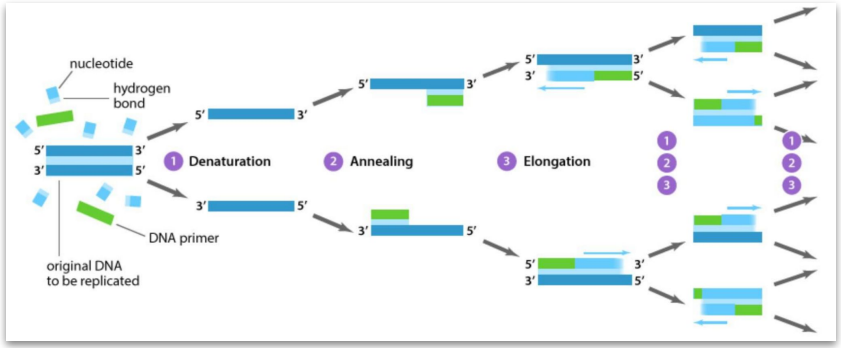
Líneas con la mutación G12C

Show 10 entries — Export: CSV TSV — Search:

Sample name	Gene name	Transcript	Primary Tissue	Tissue Subtype 1	Primary Histology	Histology Subtype 1	Pubmed ID	Zygosity	Somatic Status	Sample Type
	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Adenocarcinoma	-	Homozygous	Unknown	Cultured
	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Adenocarcinoma	-	Homozygous	Unknown	Cultured
	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Non small cell carcinoma	-	Homozygous	Unknown	Cultured
	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Non small cell carcinoma	-	Heterozygous	Unknown	Cultured
NCI-H358	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Bronchioloalveolar adenocarcinoma	-	Heterozygous	Unknown	Cultured
	KRAS	ENST00000256078.8	Lung	NS	Carcinoma	Non small cell carcinoma	-	Homozygous	Unknown	Cultured
KYSE-410	KRAS	ENST00000256078.8	Oesophagus	NS	Carcinoma	NS	-	Heterozygous	Unknown	Cultured
OV-56	KRAS	ENST00000256078.8	Ovary	NS	Carcinoma	NS	-	Heterozygous	Unknown	Cultured
MIA-PaCa-2	KRAS	ENST00000256078.8	Pancreas	NS	Carcinoma	Ductal carcinoma	-	Homozygous	Unknown	Cultured
UM-UC-3	KRAS	ENST00000256078.8	Urinary tract	Bladder	Carcinoma	NS	-	Homozygous	Unknown	Cultured

Showing 11 to 20 of 20 entries — First Previous 1 2 Next Last

¿Cómo haría una PCR de este gen?



Respuesta: Primer-BLAST

Enlace

Pregunta 6

Secuencia a amplificar

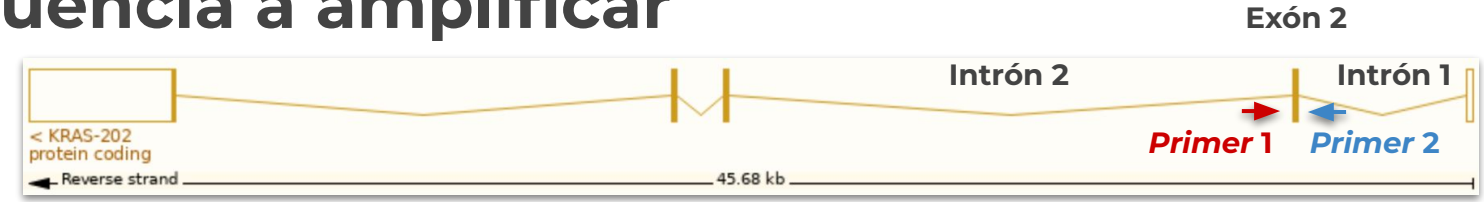
En la sección *Sequence* click en *Exons*

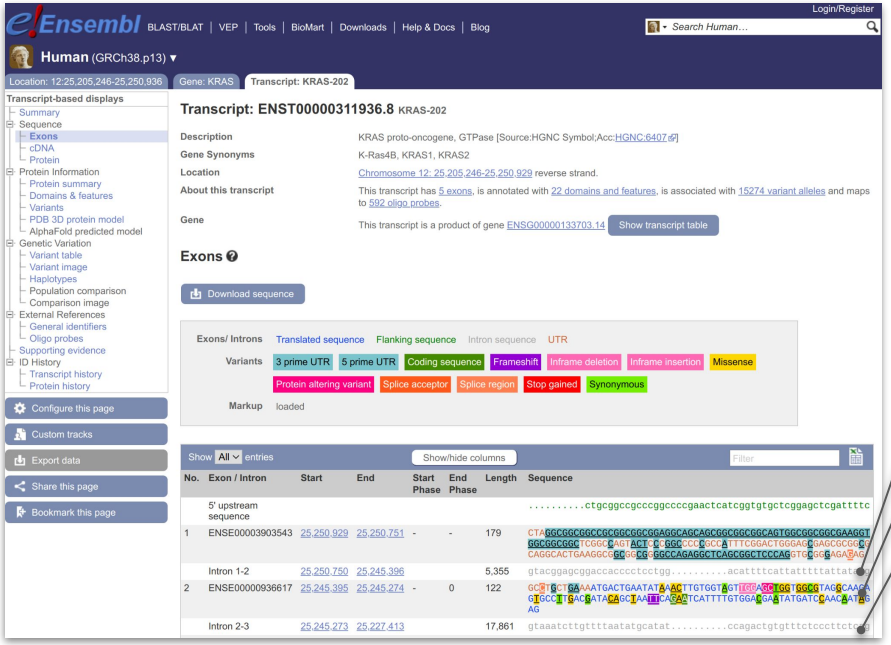
Leyenda:

- UTRs
- Exones codificantes
- Intrones

En la Práctica 2 queremos amplificar el exón 2
(el que está mutado)

Secuencia a amplificar





Secuencia del intrón 1 (incompleta)

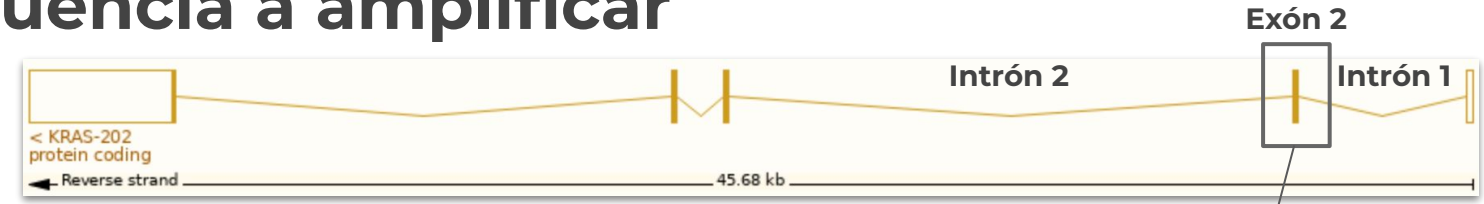
Secuencia del exón 2

Secuencia del intrón 2 (incompleta)

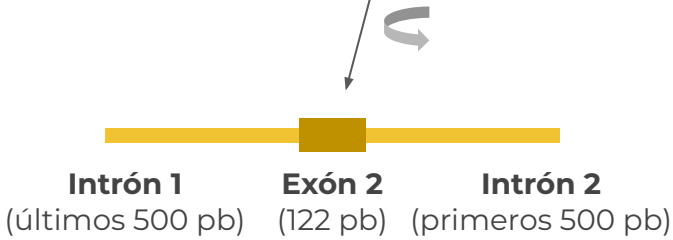
Las secuencias completas se descargan en el botón "**Download sequence**", pero nosotros vamos a usar la que está subida en **PRADO**:

"Secuencia de KRAS para el diseño de primers"

Secuencia a amplificar




```
>KRAS intron 1 last 500bp - exon 2 - intron 2 first 500bp
TTAAAAGTTTTTCTGTAGCTGTTGCATATTGACTTCTAACACTTAGAGGTGGGGGTCCAC
TAGGAAAACTGTAACAATAAGAGTGGAGATAGCTGTCAGCAACTTTTGTGAGGGTGTGCT
ACAGGGTGTAGAGCACTGTGAAGTCTCTACATGAGTGAAGTCATGATATGATCCTTTGAG
AGCCTTTAGCCGCCGCAGAACAGCAGTCTGGCTATTTAGATAGAACAACTTGATTTTAAG
ATAAAAGAACTGTCTATGTAGCATTTATGCATTTTTCTTAAGCGTCGATGGAGGAGTTTG
TAAATGAAGTACAGTTCATTACGATACACGTCTGCAGTCAACTGGAATTTTCATGATTGA
ATTTTGTAAGGTATTTTGAAATAATTTTTCATATAAAGGTGAGTTTGTATTAAAAGGTAC
TGGTGGAGTATTTGATAGTGTATTAACCTTATGTGTGACATGTTCTAATATAGTCACATT
TTCATTATTTTTATTATAAGGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGG
AGCTGGTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGA
CGAATATGATCCAACAATAGAGGTAAATCTTGTTTTAATATGCATATTACTGGTGCAGGA
CCATTCTTTGATACAGATAAAGGTTTCTCTGACCATTTTCATGAGTACTTATTACAAGAT
AATTATGCTGAAAGTTAAGTTATCTGAAATGTACCTTGGGTTTCAAGTTATATGTAACCA
TTAATATGGGAACTTTACTTTCCTTGGGAGTATGTCAGGGTCCATGATGTTCACTCTCTG
TGCATTTTGATTGGAAGTGTATTTCAGAGTTTCGTGAGAGGGTAGAAATTTGTATCCTAT
CTGGACCTAAAAGACAATCTTTTTATTGTAACTTTTATTTTTATGGGTTTCTTGGTATTG
TGACATCATATGTAAAGGTTAGATTTAATTGTACTAGTGAAATATAATTGTTTGATGGTT
GATTTTTTTAAACTTCATCAGCAGTATTTTCCTATCTTCTTCTCAACATTAGAGAACCTA
CAACTACCGGATAAATTTTACAAAATGAATTATTTGCCTAAG
```



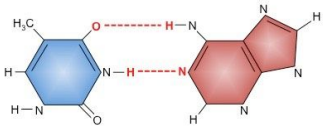
El archivo está en **formato FASTA**:

Formato para representar secuencias de nucleótidos y aminoácidos.

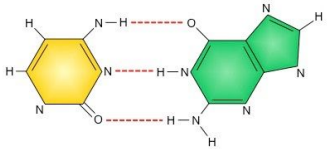
- Línea de **encabezado** que comienza por “>”. Contiene información sobre la secuencia (ID, nombre…)
- Líneas de **secuencia**

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):**
  - 50-60°C
  - Menos de 5°C 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

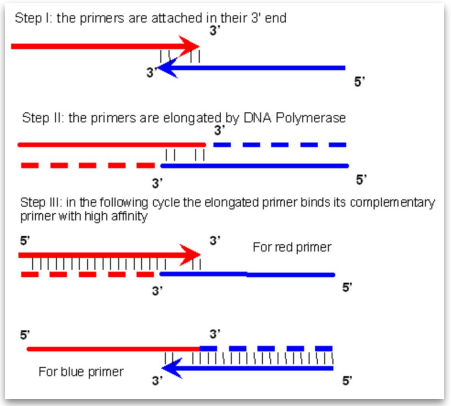


Thymine / Uracil pairs with Adenine
(2 hydrogen bonds)



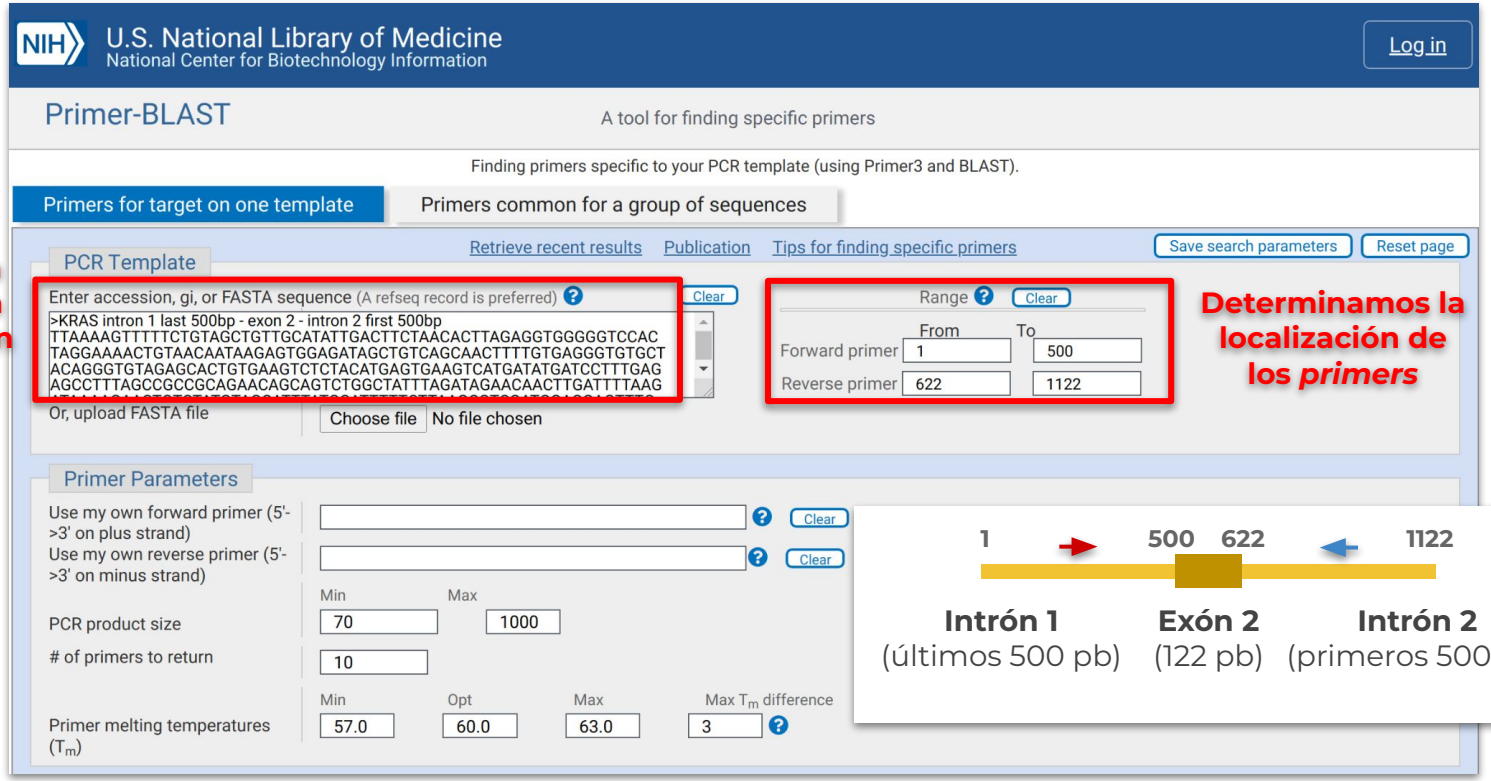
Cytosine pairs with Guanine
(3 hydrogen bonds)

Formación de dímeros de primers



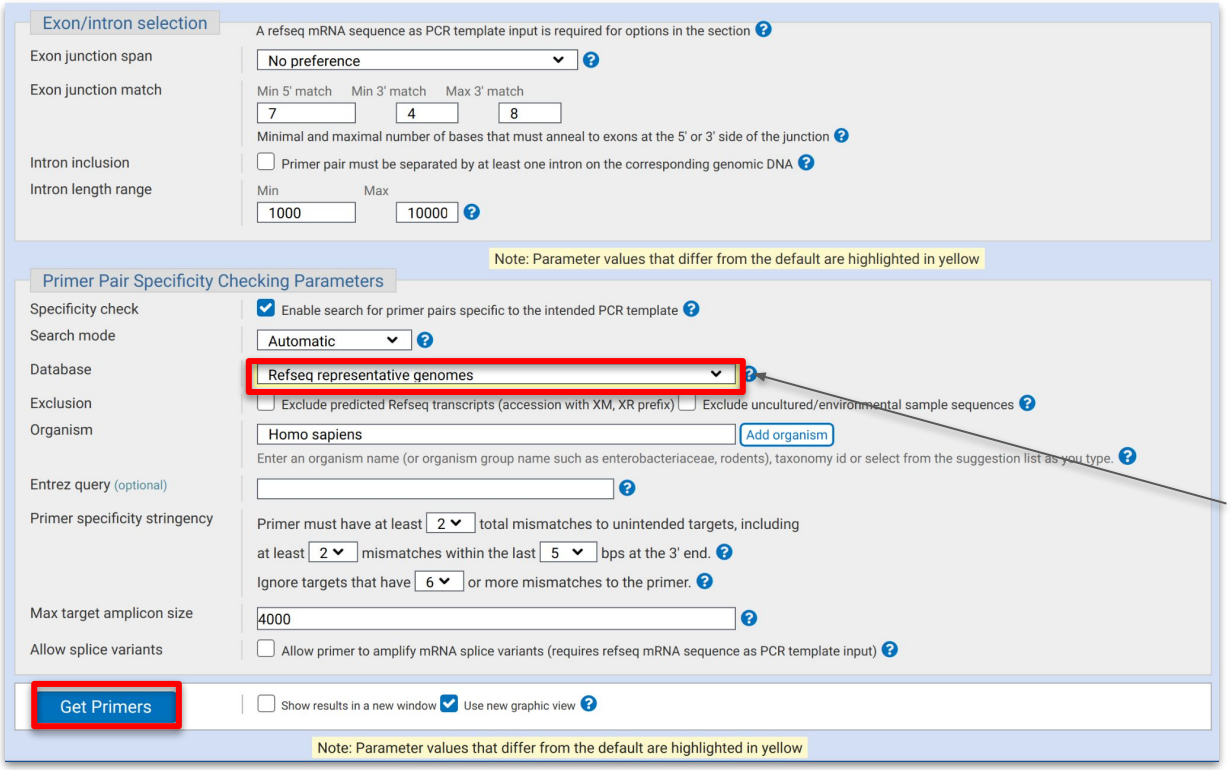
Primer-Blast

Pegamos la secuencia a amplificar en formato FASTA



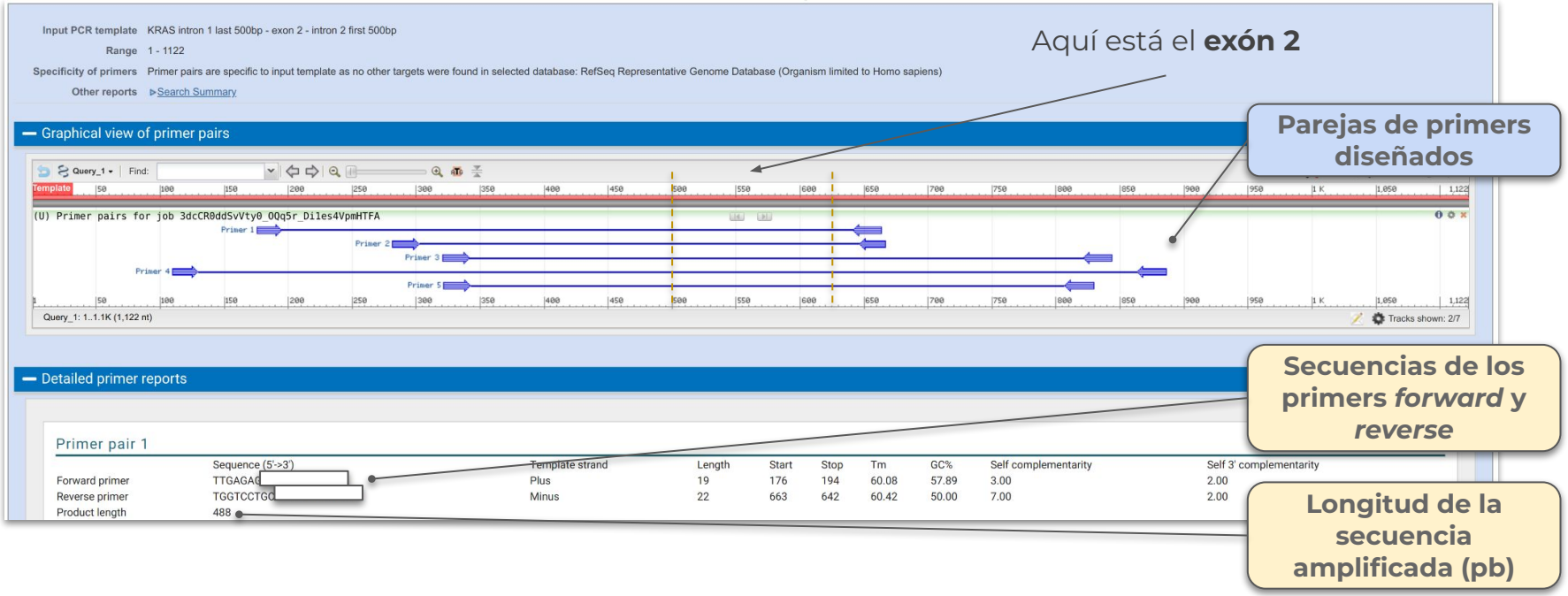
Determinamos la localización de los *primers*

Primer-Blast



Primer-Blast tiene en cuenta la secuencia del **resto del genoma humano** para no generar *primers* con hibridaciones **inespecíficas**

Primer-Blast: resultados



Entrega en PRADO:
hasta 22/03/2024 - 14:00h

- Si ya has contestado las 6 preguntas, súbelas a la entrega de PRADO.
- Si necesitas más tiempo, la entrega estará abierta hasta el **viernes 22 de marzo de 2024 a las 14:00**.

CUESTIONES SEMINARIO 1:
HERRAMIENTAS BIOINFORMÁTICAS EN GENÓMICA
Grado en Medicina | Biología Molecular | Curso 2023/24

Nombre y Apellidos:

1. Sobre el gen *KRAS*:
   - **a)** ¿Cuál es su Ensembl ID?
   - **b)** ¿En qué cromosoma, coordenadas y hebra se localiza?
   - **c)** Nombra sus 2 genes más próximos.
   - **d)** Indica el dbSNP ID y la consecuencia de una variante patogénica de este gen.
2. Sobre la isoforma (transcrito) de referencia del gen *KRAS*:
   - **a)** ¿Cuál es su Ensembl ID?
   - **b)** ¿Cuál es su longitud en pares de bases y en aminoácidos?
   - **c)** ¿Cuántos exones tiene?
3. Indica el identificador y la secuencia de un miRNA maduro que regule la expresión del gen *KRAS*.
4. ¿Qué laboratorio en España realiza el diagnóstico genético de mutaciones en KRAS?
5. Nombra otra línea celular distinta a NCI-H358 que contenga la mutación G12C en KRAS.
6. Indica la pareja de primers que has diseñado para amplificar el exón 2 completo de KRAS.