



# Identificación de mutaciones patogénicas.

**Prácticas de Genética Médica.  
1º Curso. Grado en Medicina.**

**Departamento de Biología, Genética, Histología y  
Farmacología.**

**Proyecto de innovación docente. PDI\_003\_2021/2022.**

Desarrollo de un repositorio de datos genómicos e implantación de prácticas online para mejorar el aprendizaje en el análisis genómico humano.



**Universidad de Valladolid**

# 1. Introducción.

## 1.1 Identificación de mutaciones patogénicas.

La identificación de variantes patogénicas es uno de los principales objetivos en el diagnóstico genético, ya que permite conocer la base genética de una enfermedad y, por lo tanto, la toma de decisiones sobre el tratamiento, la evolución y pronóstico de dicha patología.

Aunque existen diferentes estrategias para la detección de variantes genéticas, la secuenciación bidireccional de exones y zonas de splicing mediante el método Sanger es una de las técnicas más utilizadas en los laboratorios clínicos. Para realizar esta técnica se requiere una muestra de sangre periférica u otro tejido del individuo, a partir de la cual se aísla su ADN genómico. A continuación, mediante la reacción en cadena de la polimerasa (PCR) se amplifican todos los exones, junto con las secuencias intrónicas flanqueantes de cada exón que son examinados nucleótido a nucleótido mediante la reacción de secuenciación. De esta manera se puede identificar cualquier diferencia con respecto a la secuencia de referencia de dicho gen (podéis recordar la técnica visualizando el siguiente video: <https://www.youtube.com/watch?v=NEu0mO-2ras>).

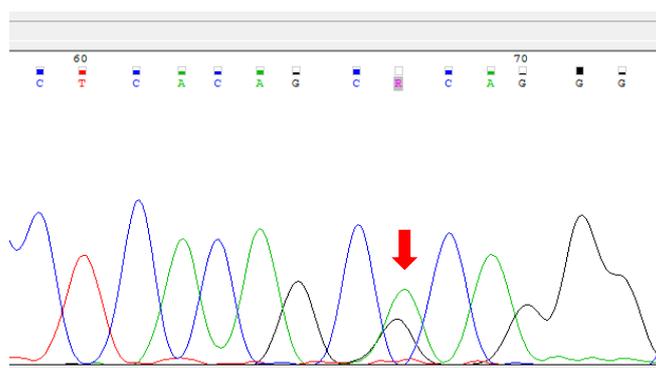


Figura 1: Electroferograma procedente de una paciente con una mutación puntual en heterocigosis.

## 1.2 Tipos de mutaciones.

Mediante la secuenciación del ADN se pueden identificar distintos tipos de variantes de secuencia que alteran a un único nucleótido (puntuales) o a unos pocos. Las variantes de secuencia se pueden clasificar como **mutaciones claramente patogénicas**, si truncan la proteína resultante o si alteran las secuencias canónicas de splicing (AG/GT). En este grupo se hallan las **mutaciones denominadas sin sentido (nonsense)** que originan un codón de stop (UAG, UAA o UGA), que termina prematuramente la traducción produciendo una proteína truncada. También las mutaciones denominadas de **cambio de pauta de lectura (frameshift)** que implican la **inserción o la delección de uno o varios nucleótidos cuyo número no es un múltiplo de tres**, de forma que la pauta de lectura de la secuencia del ARNm (ARN mensajero) queda alterada. El resultado es una proteína truncada, debido a la creación de un codón de stop más adelante. Por último, las mutaciones de splicing que alteran el procesamiento del ARNm en la eliminación de los intrones y unión de los exones. Las mutaciones de splicing pueden producirse en distintas partes de un gen, pero las que se consideran claramente patogénicas son las que modifican las

secuencias canónicas donadoras (GT) o aceptadoras (AG) de splicing situadas en los dos nucleótidos iniciales y finales de cada intrón.

Por otro lado, algunas variantes de secuencia tienen una repercusión funcional incierta en la proteína resultante puesto que **no alteran su pauta de lectura (in frame)**. Estos cambios a priori son considerados **variantes de secuencia no clasificadas (UCVs, unclassified sequence variants)**. En este grupo se incluyen las variantes de cambio de aminoácido denominadas de **cambio de sentido (missense)**, **deleciones o inserciones de un número de nucleótidos múltiplo de tres** y las **variantes de secuencia potencialmente implicadas en el splicing** pero que no alteran los nucleótidos canónicos (AG/GT).

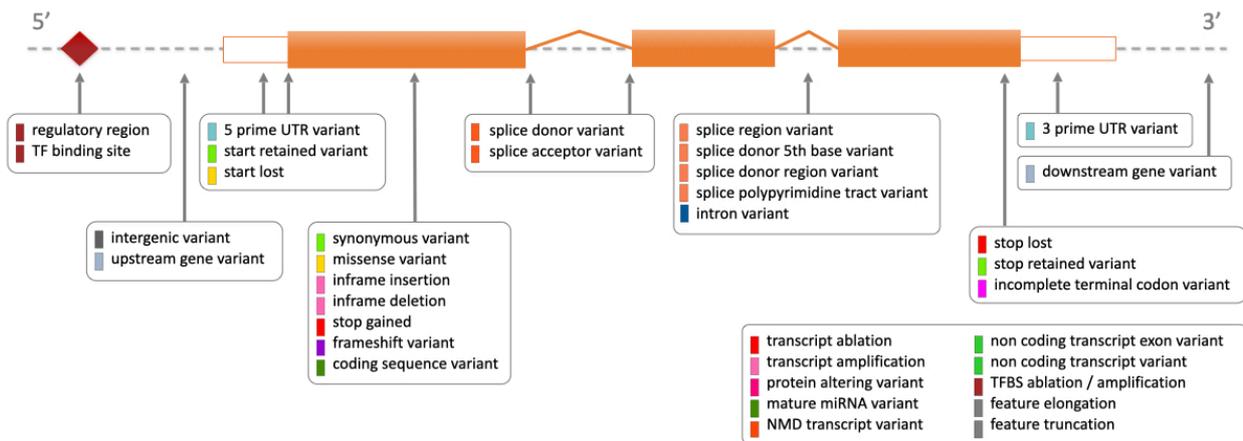


Figura 1: Tipos de variantes genéticas. Fuente: NCBI.

### 1.3 Evaluación del impacto clínico de las variantes genéticas.

Como ya se ha comentado, las variantes de secuencia que truncan la proteína resultante o que alteran las secuencias canónicas de splicing son claramente patogénicas. El problema surge cuando se detecta una variante de significado incierto. La mejor forma de evaluar la patogenicidad de este tipo de variantes es la realización de un análisis funcional, pero no siempre es posible, dado que se trata de estrategias muy laboriosas no siempre al alcance un de un laboratorio de análisis clínico. Para ello se recurre a otras estrategias que utilizan herramientas “in silico” para evaluar el impacto de la variante:

- 1- La búsqueda de datos en bases de datos de mutaciones, polimorfismos, y/o datos bibliográficos.
- 2- El grado de conservación del aminoácido en un alineamiento múltiple de secuencias con proteínas ortólogas.
- 3- La diferencia bioquímica y biofísica entre el aminoácido original y el mutado.
- 4- Programas de predicción de patogenicidad (Polyphen y SIFT)
- 5- Análisis de la segregación de la variante con la enfermedad en la familia.
- 6- El análisis de cromosomas control.

Estos análisis permiten clasificar las variantes UCVs como: 1) muy probablemente patogénica, 2) probablemente patogénica, 3) variante de efecto indeterminado y 4) probablemente neutra (polimorfismo).

## 2. Objetivo de la práctica:

El objetivo de esta práctica es aprender a identificar variaciones de secuencias obtenidas por el método de secuenciación SANGER, así como a interpretar el impacto de la variante identificada en la función de la proteína y en la salud del paciente.

## 3. Metodología:

Para realizar esta práctica necesitamos:

- Ordenador con conexión a internet.
- Programa de análisis de secuencia: Chromas o Finch TV.
- Electroferogramas Sanger procedentes de casos reales (repositorio online en campus virtual).
- Bases de datos y herramientas de análisis online:
  - Ensembl: <https://www.ensembl.org/index.html>
  - ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>
  - PolyPhen: <http://genetics.bwh.harvard.edu/pph2/>

### Nota: Instalación y manejo del programa: Finch TV.

Accede al campus virtual y en la carpeta “práctica 13: Identificación de variantes” encontrarás una carpeta comprimida con el nombre FINCHTV. Accede e instala el programa FINCHTV. Visualiza el video “Manejo FINCHTV” disponible en la misma carpeta para saber como se maneja el programa.

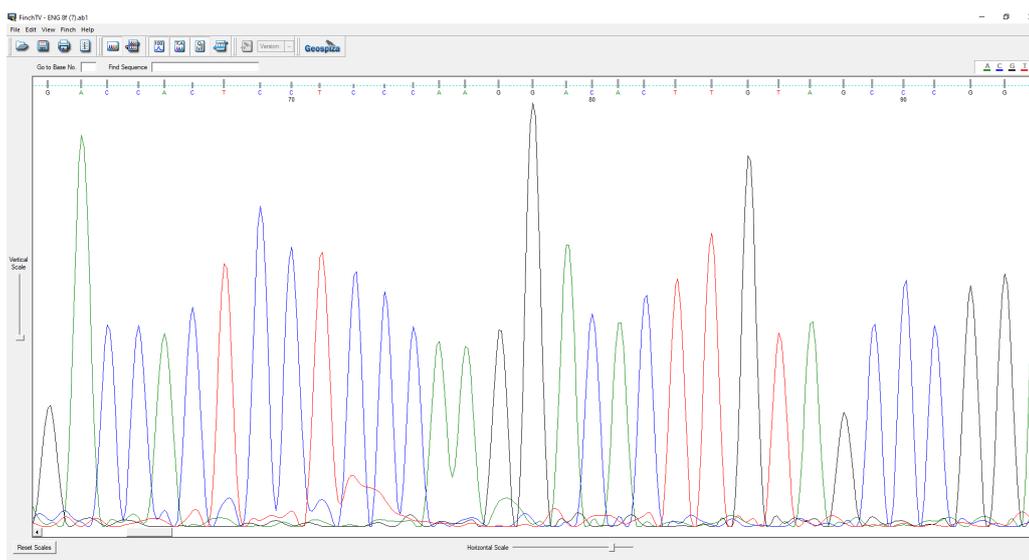


Figura 2: Programa FINCHTV

## 4. Desarrollo de la práctica:

En la carpeta: “práctica 13: Identificación de variantes” podrás encontrar una serie de archivos con formato “abi” que corresponden a los electroferogramas que vamos a explorar en esta práctica. Como se ha explicado en la introducción, con este tipo de tecnología secuenciamos tanto la hebra de ADN forward (+) como la hebra de ADN reverse (-), por eso cada paciente tendrá mínimo dos electroferogramas. Hay que comprobar ambas secuencias de cada paciente que vayamos a analizar.

En esta práctica vamos a analizar dos secuencias ejemplo que corresponden a los pacientes 1 y 2. Después dispondrás de un repositorio de secuencias de diferentes pacientes para practicar de forma autónoma desde casa.

### Paciente 1

#### 4.1 Identificación de variantes.

**Paso I.** Abre el archivo ENG\_48\_fw con el programa FINCHTV e identifica la alteración:

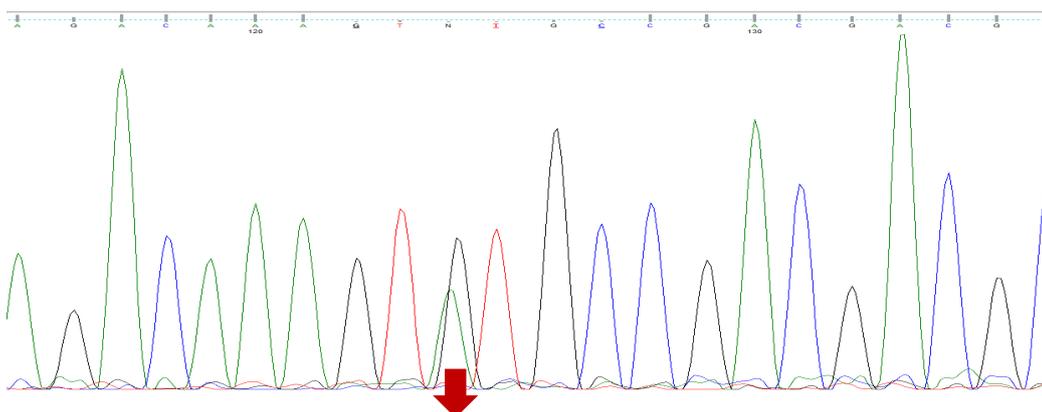


Figura 3: Variante identificada.

*En el ejemplo puedes observar una posición donde vemos dos picos que solapan: uno negro (que corresponde a una G) y uno verde (que corresponde a una A). Se trata de una sustitución de una base por otra en heterocigosis.*

- **Paso II.** Una vez identificada la variante, desde el mismo programa, podemos comparar la secuencia que estamos analizando con una secuencia de referencia, mediante el análisis de alineamiento de secuencias “BLAST”. Esto nos permitirá determinar el gen que estamos analizando (en base a la similitud de secuencias) y por otro lado la variante que hemos encontrado (es decir, el cambio de nucleótidos/nucleótidos presentes en nuestra secuencia). Para ello, en el propio FINCHTV seleccionamos con el ratón un fragmento de la secuencia de interés que incluya la variante (unos 50-100 nucleótidos) o si la variación es muy grande, la zona

inmediatamente anterior o posterior a la variante. Pinchamos con el botón derecho para que aparezca un desplegable en el que seleccionaremos la opción: “BLAST Sequence”- “Nucleotide, BLASTn”; tal y como se muestra en la figura.

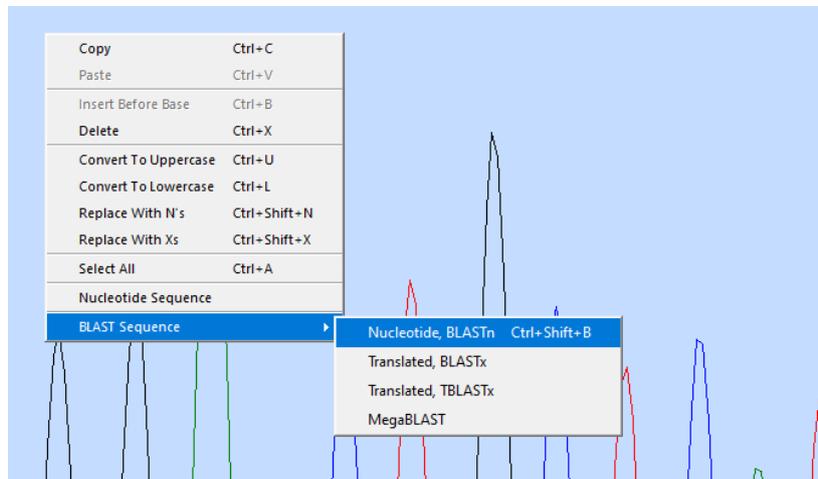


Figura 4: Análisis de alineamiento BLAST desde FINCHTV.

Automáticamente el programa nos redirecciona a la página web del NIH (National Library of Medicine) donde podremos hacer el BLAST de nuestro fragmento de secuencia. Dejamos la configuración que viene por defecto y pinchamos sobre el botón azul (BLAST). Esperamos a que el programa haga su búsqueda.

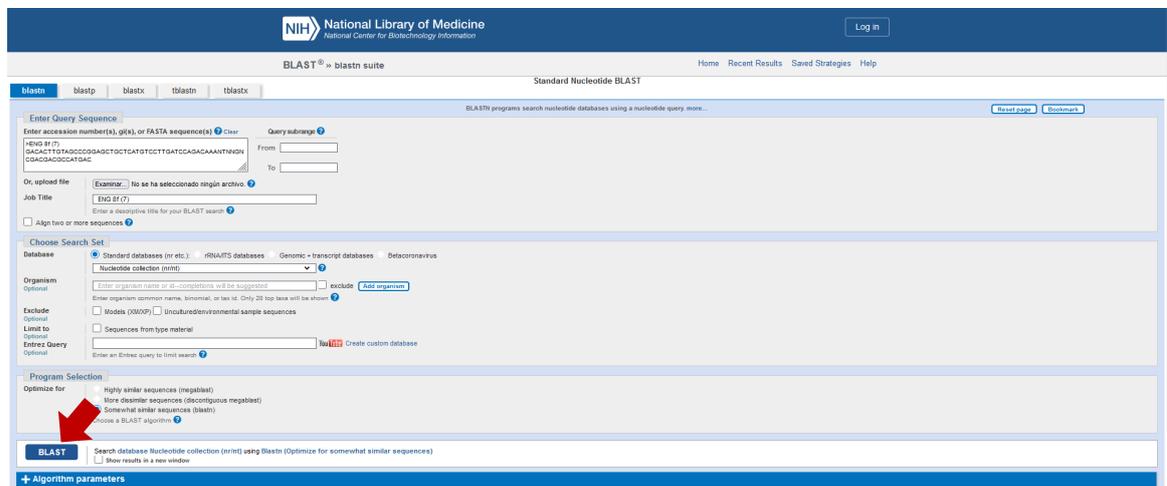


Figura 5: National Library of Medicine: BLASTn.

- **Paso III:** Una vez finaliza la búsqueda, el programa nos mostrará todas las secuencias que muestran mayor similitud con la secuencia que hemos introducido. Seleccionamos la que tenga un mayor grado de similitud (cobertura: Query cover). Ten en cuenta que las muestras que estamos analizando son muestras de pacientes, por lo que el organismo de referencia debe ser: “Homo sapiens”.

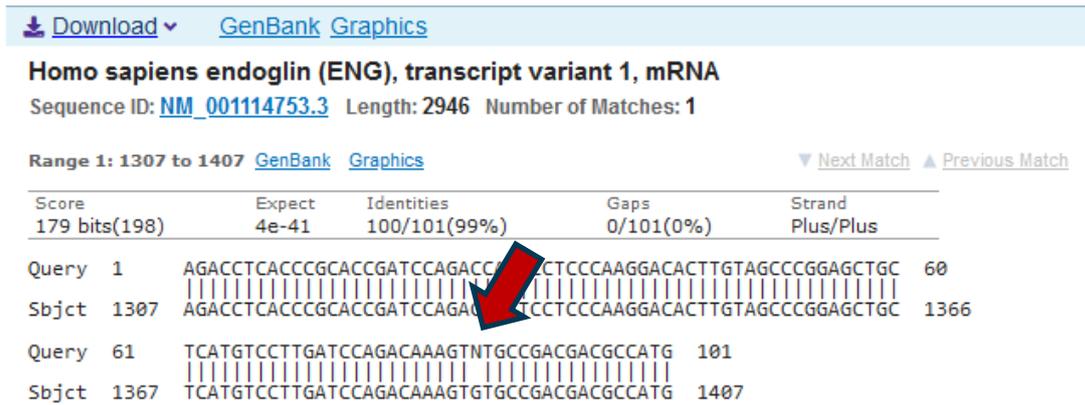


Figura 6: Resultados del BLASTn. En nuestro ejemplo: pinchariamos sobre la primera secuencia que corresponde con el ARNm del gen ENG de humano.

Como puedes ver en el ejemplo, la secuencia que estamos analizando (query) corresponde al transcrito del gen de la endoglina, que presenta una similitud del 99% con la secuencia de referencia (sbjct). Esto se debe a que de 101 nucleotidos que hemos introducido en el programa, 1 es diferente a la secuencia de referencia. En nuestra secuencia este cambio de base esta marcado con una N, mientras que en la secuencia de referencia la base nitrogenada que conforma el nucleótido sería una guanina (G). Si vuelves al programa FINCHTV puedes ver que en este caso el cambio que se ha producido es de una G por una A (esto se representa de la siguiente manera: G>A).

- **Paso IV:** Comprueba que la variante identificada se encuentra tanto en la hebra forward como en la hebra reverse (para ello no olvides darle la vuelta a la secuencia reverse).

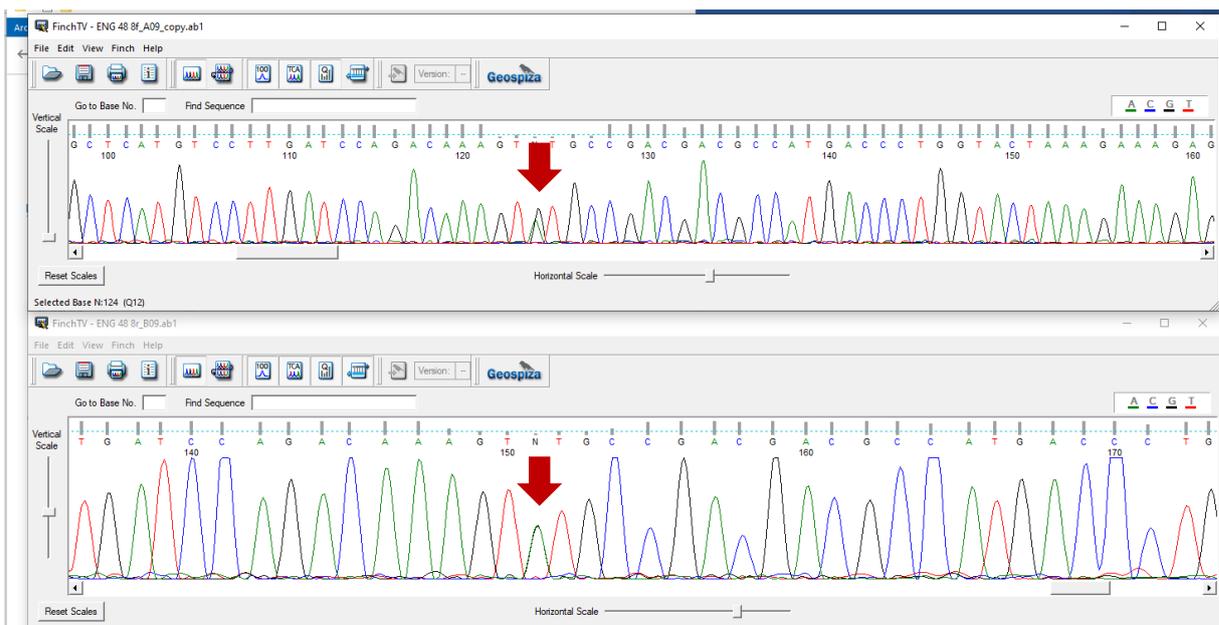


Figura 7: Comparación de la secuencia forward y reverse. Fijate como en la secuencia reverse los picos se superponen tanto que es difícil identificar la variante.

## 4.2 Evaluación del impacto de la variante en la estructura y función de la proteína.

Una vez identificado el gen de estudio y la variante de interés, debemos averiguar que impacto tendrá la variante en la estructura y función de la proteína y por lo tanto en la salud del paciente. Para ello:

- **Paso I.** Abrimos la página web: ENSEMBL: <https://www.ensembl.org/index.html>. Introducimos el nombre del gen y la especie que estamos analizando.

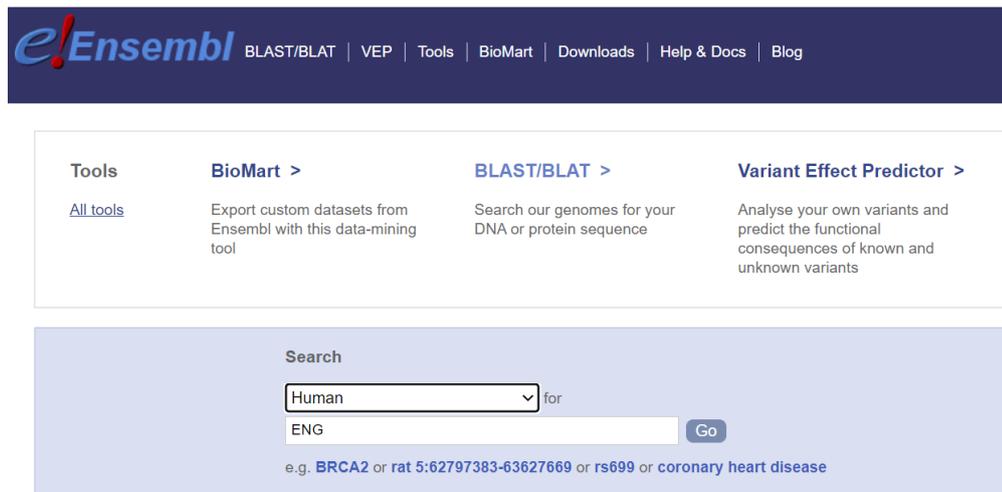


Figura 8: Consulta en la base de datos ensembl.

- **Paso II.** De todas opciones que nos proporciona ensembl seleccionamos el gen de interes que exploraremos.

*De nuevo tomando como ejemplo el gen de la endoglina (ENG), obtenemos un listado genes. En este caso pinchamos sobre la primera entrada de la lista.*

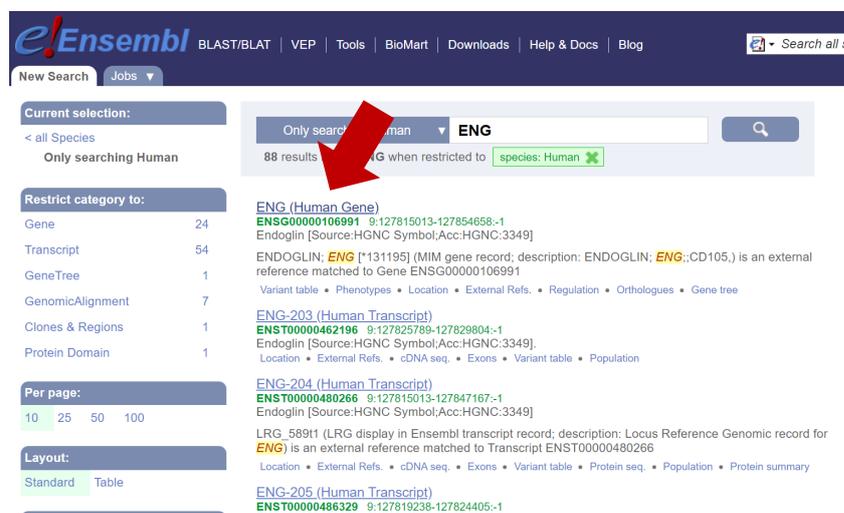


Figura 9: Resultados de una búsqueda realizada con ensembl.

**Paso III.** Al pinchar sobre el gen aparece la descripción del mismo, el número de transcritos y mucha información sobre la secuencia. En este punto nos centraremos en la tabla de transcritos para seleccionar el transcrito con más información disponible para explorarlo. Es importante que te fijas en los códigos que usa Ensembl para referirse a un gen (**ENSG**), un transcrito (**ENST**) o una proteína (**ENSP**).

Pincha sobre el ENST que disponga de más información (por ejemplo, un transcrito que contenga una proteína codificante, de mayor tamaño, con código UniProt y RefSeq Match y que presente muchos “flags”, que identifican transcritos de mayor calidad o más relevantes).

**Gene: ENG** ENSG00000106991

Description: endoglin [Source:HGNC Symbol;Acc:HGNC:3349]

Gene Synonyms: CD105, END, HHT1, ORW, ORW1

Location: Chromosome 9: 127,815,013-127,854,658 reverse strand. GRCh38:CM000671.2

About this gene: This gene has 5 transcripts (splice variants), 169 orthologues, 2 paralogues and is associated with 5 phenotypes.

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	Flags
ENST00000373203.9	ENG-202	2946	658aa	Protein coding	CCDS48029.6	P17813-1	NM_001114753.3	MANE Select v0.95 Ensembl Canonical GENCODE basic APPRIS P4 TSL:1
ENST00000344849.4	ENG-201	3059	625aa	Protein coding	CCDS6880.6	P17813-2	-	GENCODE basic APPRIS ALT2 TSL:1
ENST00000480266.5	ENG-204	2804	476aa	Protein coding	CCDS75906.6	F5G388.6	-	GENCODE basic TSL:2
ENST00000486329.1	ENG-205	663	No protein	Processed transcript	-	-	-	TSL:2
ENST00000462196.1	ENG-203	495	No protein	Processed transcript	-	-	-	TSL:3

**Summary**

Name: ENG (HGNC Symbol)

CCDS: This gene is a member of the Human CCDS set: CCDS48029.1, CCDS6880.1, CCDS75906.1

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: P17813

RefSeq: This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table

LRG: LRG\_589 provides a stable genomic reference framework for describing sequence variants for this gene

Ensembl version: ENSG00000106991.14

Other assemblies: This gene maps to 130,577,292-130,616,937 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000106991

Gene type: Protein coding

Annotation method: Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article

Figura 10: Información del gen. En este ejemplo pinchamos sobre el primer transcrito de la lista.

- **Paso IV.** Una vez seleccionamos el transcrito que nos interesa aparecerá en el menú de la parte izquierda un desplegable que nos permite explorar los exones, el cDNA y la proteína codificante.

**Human (GRCh38.p13)**

Location: 9:127,815,013-127,854,658 Gene: ENG Transcript: ENG-202 Jobs

**Transcript: ENST00000373203.9** ENG-202

Description: endoglin [Source:HGNC Symbol;Acc:HGNC:3349]

Gene Synonyms: CD105, END, HHT1, ORW, ORW1

Transcript-based displays: Summary, Sequence, Exons, cDNA, Protein

Figura 11. Menú para explorar el transcrito del gen de interés seleccionado.

- Si **seleccionamos la opción “exones”** podemos hacer una búsqueda del fragmento de la secuencia donde se encontraba nuestra variante. De esta manera podremos ver en qué exón se localiza nuestra variante y si ya se ha descrito con anterioridad. Para ello, copiamos un fragmento pequeño de unos 10-15 nucleótidos de la secuencia inmediatamente anterior o posterior a la variante que hemos identificado y la buscamos (Ctrl + F) en la secuencia de nucleótidos de los exones que muestra Ensembl.

Ten en cuenta que la secuencia de nucleótidos tiene varias líneas por lo que la variación puede estar al inicio o final de la secuencia y puede resultar difícil identificar el fragmento.

The image shows the Ensembl genome browser interface. At the top, there are navigation options like 'Transcript history', 'Configure this page', 'Custom tracks', 'Export data', 'Share this page', and 'Bookmark this page'. Below this is the 'Exons' section, which includes a 'Download sequence' button and a list of variants with categories like 'Promoter altering variant', 'Splice acceptor', 'Splice donor', 'Splice region', 'Start codon', 'Stop gained', 'Stop lost', 'Stop retained', and 'Synonymous'. A table below lists 9 exons with columns for 'No.', 'Exon / Intron', 'Start', 'End', 'Start Phase', 'End Phase', 'Length', and 'Sequence'. The 'Sequence' column shows a multi-line representation of the nucleotide sequence. Below the table, a detailed view of the sequence is shown, with a red arrow pointing to a specific variant: TTTGATCCAGACAAAGT. The sequence is color-coded by nucleotide type (A, C, G, T) and the variant is highlighted in red.

Figura 12. Búsqueda de la variante en los exones del transcrito seleccionado. (ejemplo: TTTGATCCAGACAAAGT). En nuestro ejemplo la variante identificada se localiza en el exón 7.

**Paso V.** Una vez identificado el nucleótido afectado, podemos explorar las variantes descritas. Como habrás podido comprobar, cada nucleotido tiene un color/colores distinto que hace referencia al tipo de mutación o variante que se ha descrito en esa posición.

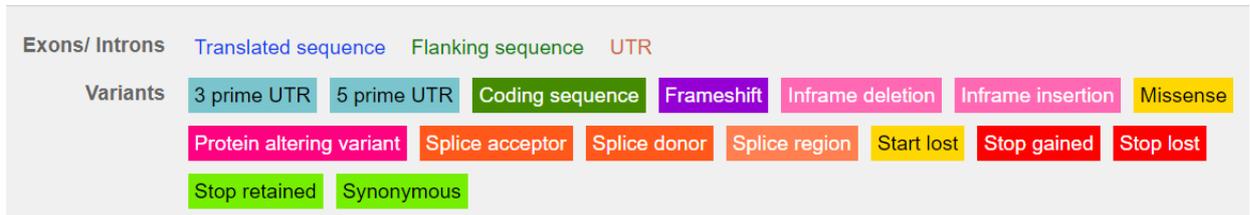


Figura 13. Código de colores que identifican el tipo de variante encontrada.

En nuestro ejemplo el nucleótido G donde se localiza la variante (G>A) aparece en amarillo donde se ha descrito una mutación “missense”.

**Paso VI.** Podemos pinchar sobre el nucleótido y explorar qué variantes se han descrito y si alguna corresponde a la variante que hemos identificado.

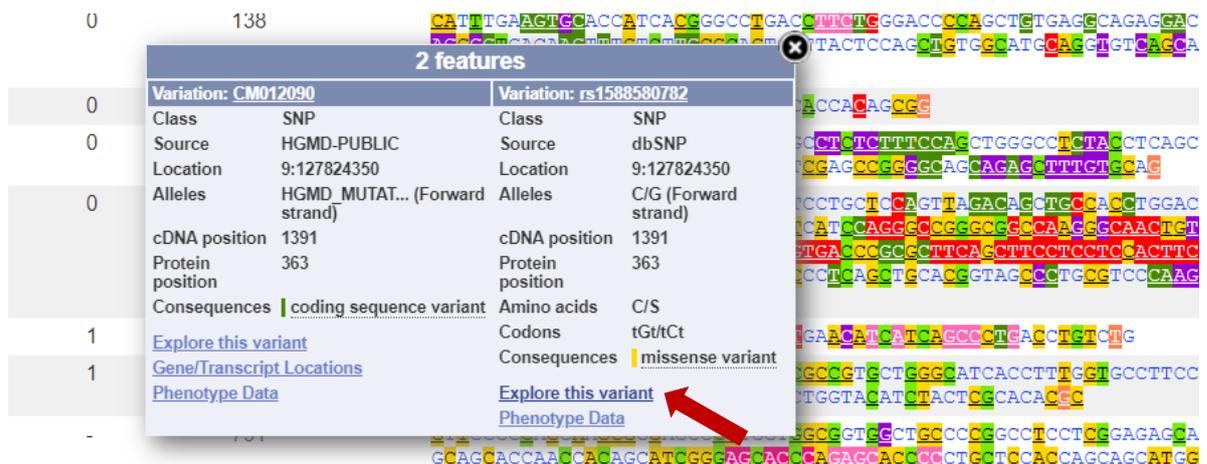


Figura 14. Variantes descritas en el nucleótido de interés.

En nuestro ejemplo, se han descrito dos variantes, aunque ninguna de ellas corresponde con la que hemos identificado (G>A).

Puesto que hay una parecida (C/G) aunque no sea la misma la vamos a explorar.

Para ello le damos a “explore this variant” y de nuevo aparece otra ventana sobre con información sobre la variante. Fíjate que las variantes descritas tienen un código que siempre se identifica como: rs seguido de un número (ej: rs1588580782).

La variante que hemos elegido es una variante de cambio de sentido (missense) y si exploramos las consecuencias fenotípicas (Phenotype Data), la variante en cuestión se relaciona con la TELANGIECTASIA HEMORRÁGICA HEREDITARIA.

A)

B)

Phenotype, disease and trait	Source(s)	Mapped Terms	Ontology Accessions	Supporting evidence	External reference	Clinical significance	Reported gene(s)	Associated allele
HEREDITARY HEMORRHAGIC TELANGIECTASIA	ClinVar [Invitae]	-	Orphanet:774	Evidence	-	★☆☆☆	ENG	G

Figura 15. A) Información sobre una variante descrita previamente. B) Consecuencias fenotípicas de la variante identificada.

**Paso VII.** Podemos explorar que consecuencia tendrá el cambio detectado en la secuencia de aminoácidos que codifica el fragmento de ADN que contiene la variante y anotar la variante. Para ello, en la pestaña de transcritos, en vez de seleccionar exones **exploraremos el cDNA.**

Copiamos de nuevo el fragmento de secuencia donde se localiza la variante e identificamos el codón que se verá afectado y el cambio que supone dicha variante.

**A)**

```

    *R***M*M *SB* BRVNVRY** * YSS*RR*** ***** ** Y*VW
1381 CAGACAAAGTGTGCCGACGACGCCATGACCCCTGGTACTAAAGAAAGAGCTTGTGCGCAT 1440
1078 CAGACAAAGTGTGCCGACGACGCCATGACCCCTGGTACTAAAGAAAGAGCTTGTGCGCAT 1137
    360 -Q--T--K--C--A--D--D--A--M--T--L--V--L--K--K--E--L--V--A--H-- 379
    
```

**B)**



**Códigos de Aminoácidos**

PRIMERA LETRA DEL CODÓN	SEGUNDA LETRA DEL CODÓN							TERCERA LETRA DEL CODÓN	
	U	C	A	G	U	C	A		G
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	STOP	UGA	STOP	A
	UUG	Leu	UCG	Ser	UAG	STOP	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

Aminoácido	Código de tres letras	Código o una letra	Aminoácido	Código de tres letras	Código o una letra
Alanina	Ala	A	Leucina	Leu	L
Arginina	Arg	R	Lisina	Lys	K
Asparagina	Asn	N	Metionina	Met	M
Aspártico	Asp	D	Fenilalanina	Phe	F
Cisteína	Cys	C	Prolina	Pro	P
Glutámico	Glu	E	Serina	Ser	S
Glutamina	Gln	Q	Treonina	Thr	T
Glicina	Gly	G	Triptófano	Trp	W
Histidina	His	H	Tirosina	Tyr	Y
Isoleucina	Ile	I	Valina	Val	V
Selenocisteína	Sec	U	Pirrolisina	Pro	O

Figura 16. A) Posición del nucleótido y el aminoácido alterado. B) Código genético.

La variante que hemos identificado (G>A) supondría un cambio en el aminoácido 363, que cambia de cisteína (C, Cys, codón: UGU) a tirosina (Y, Tyr, codón: UAU). Por lo tanto, la variante que nosotros hemos identificado es también una missense. La mutación se anota de la siguiente manera: **c.1088 G>A p.Cys363Tyr.** (donde la c.1088 hace referencia a la posición del nucleótido del cDNA alterado, y p.363 la posición del aminoácido alterado).

**Paso VIII.** Otra opción para saber si la variante es patogénica o no es consultar en diferentes bases de datos. Para ello entramos en ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>):

- Si disponemos del código rs de la variante podemos buscarlo directamente.
  - Ej.: rs1588580782: descrita como probablemente patogénica.
- Si no disponemos del código rs de la variante podemos buscarlo usando la anotación de la variante identificada junto con el código del transcrito NM.
  - Ej: NM\_001114753.3: c.1088 G>A p.Cys363Tyr.  
En nuestro caso no esta descrita.

### 4.3 Predicción bioinformática de los efectos potenciales de las variantes genéticas en la función y estructura de proteínas.

Si la variante identificada no se ha descrito previamente deberíamos predecir si ésta será patogénica o no. Para ello podemos acceder a simuladores “in silico” de los efectos potenciales de las variantes genética, por ejemplo, PolyPhen: <http://genetics.bwh.harvard.edu/pph2/>.

Para ello necesitamos:

- el código de la proteína: *ENSP00000362299*, o un archivo que contenga la secuencia completa de aminoácidos (en un formato especial denominado FASTA que podemos descargar de Ensembl).
- la posición donde se produce el cambio de aminoácido.
- los aminoácidos que cambian.

*En nuestro caso introducimos el código de la proteína del Ensembl: ENSP00000362299 la posición del aminoácido: 363 y la sustitución: C>Y.*

*Le damos a “Submit Query” y pasado un tiempo aparecerá nuestro resultado.*

A)

**PolyPhen-2** prediction of functional effects of human nsSNPs

Home About Help Downloads Batch query WHESs.db

**Query Data**

Protein or SNP identifier:

Protein sequence in FASTA format:

Position:

Substitution: AA<sub>1</sub> A R N D **C** E Q G H I L K M F P S T W Y V  
 AA<sub>2</sub> A R N D C E Q G H I L K M F P S T W **Y** V

Query description:

Submit Query Clear Check Status

[Display advanced query options](#)

B)

## Grid Gateway Interface

v2.2.8

[Help](#) | [Troubleshooting/FAQ](#)



---

**Service Name:** [PolyPhen-2](#)

**Session ID:**   Overwrite default

**Grid Status:**

Load	Health	Jobs:	Pending	Running
Light	88%		5	21

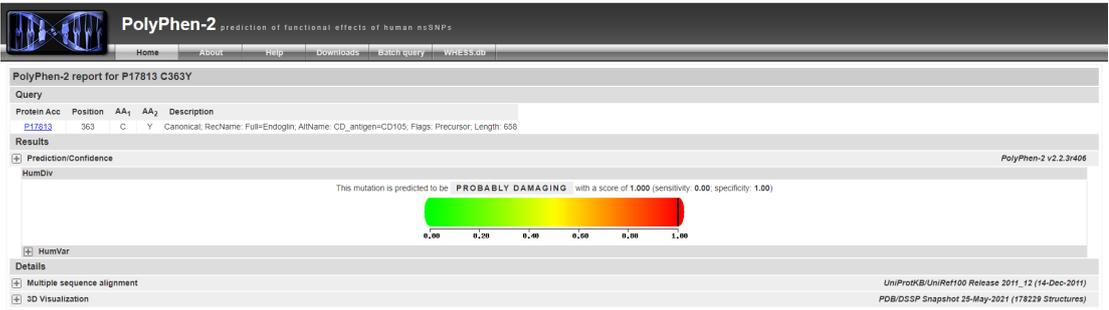
**Jobs (total):** Completed (1)

ID	Results	Errors	Date/Time	Delete	Description
8351878	<a href="#">View</a>	-	2022-05-16 13:22:11	<input type="checkbox"/>	

All items with **Delete** boxes checked will be removed!

Figura 17. A) Búsqueda en PolyPhen. B) Pantalla para acceder al resultado.

*Pinchamos en “view results”, y nos aparece el resultado: en este caso Polyphen determina que la variante identificada es probablemente patogénica.*



This mutation is predicted to be **PROBABLY DAMAGING** with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)

0.00    0.20    0.40    0.60    0.80    1.00

Figura 18. Resultado de la predicción de PolyPhen-2.

## Paciente 2:

Repetimos los pasos realizados para el paciente 1.

### 5.1. Identificación de variantes.

- **Paso I.** Abre el archivo TP53 268 4f\_E01 con el programa FINCHTV e identifica la alteración:

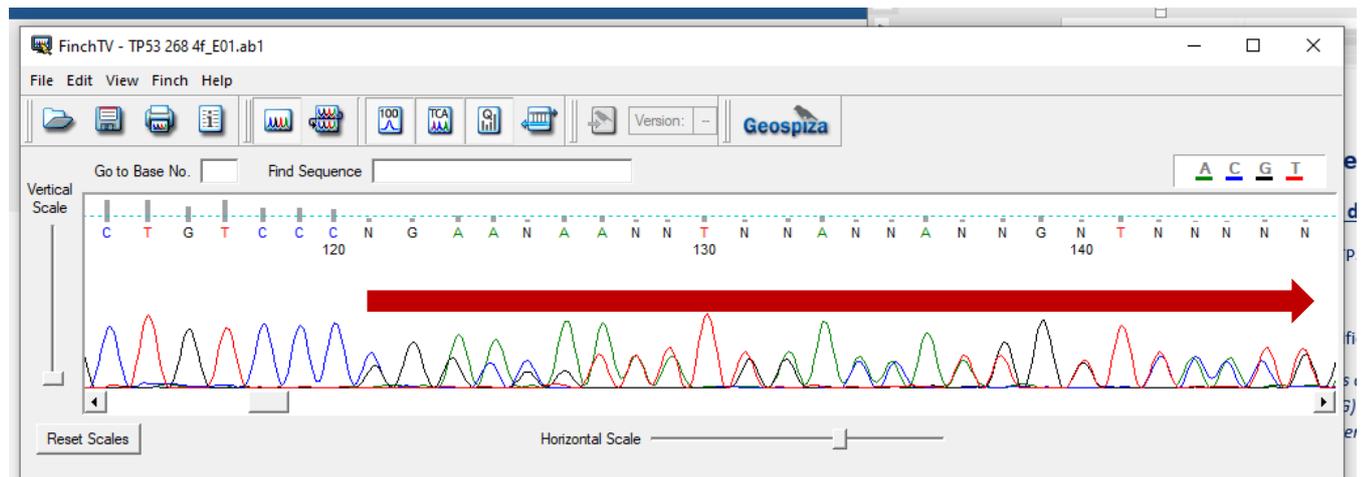


Figura 19: Variante identificada.

En el ejemplo puedes observar que apartir de una determinada posición toda la secuenica de nucleótidos cambia y por lo tanto tambien lo hara la secuenica de aminoácidos. Esto es una mutación que afecta al marco de lectura (frameshift variant).

- **Paso II y III.** Identificamos el gen y la secuenica de referencia.

Gen: TP53.

Secuenica de referencia: ATGCTGTCCC**GGACGATATTGAA**

Secuenica problema: ATGCTGTCCC**GGACGATATTGAA**

En este caso vemos claramente que se ha producido una delecion de un C que ha hecho que la secuenica se desplace a la izquierda. Esto se reproduce tanto en la hebra fordware como en la reverse. Se trata de una delecion de un nucleotido que cambia el marco de lectura y que por definición tendrá consecuencias patogénicas.

## 5.2. Impacto de la variante en la estructura y función de la proteína.

- **Pasos I-VI:** Buscamos en ensambl la variante para ver si la delección ya ha sido descrito y a que enfermedad se asocia.

*La variante se localiza en el exon 4 del gen TP53.*

*Se ha identificado un variante idéntica: rs1567557016 que se relaciona con el Síndrome de Li-Fraumeni.*

- **Paso VII y VIII:** Buscamos la posición del nucleótido afectado y del aminoácido en ensambl.

```

*VB**BYHR*YB*****R*YRR***S**R**R***H***H*****R***R**
241 CCCTTGCCTCCCAAGCAATGG GTATTGATGCTGTCCTCCGGA GATATTGACAGT 300
99 CCCCTTGCCTCCCAAGCAATGGATGATT GATGCTGTCCTCCGGA CGATATTGAACAATG 158
33 --P--L--P--S--Q--A--M--D--D--L--M--L--S--P--D--D--I--E--Q--W
53

*Y*****RH***R**YM*RWK**RB***** S**RBY**SY*****
301 GTCTCTGAGACCCGCTC GTCTCAGTCA CTCCCAATG GAGCGTCTCCCGCT 360
159 GTTCACTGAGACCCAGGTCAGATGAAGTCCCAAGT GAGCGTCTCCCGCT 218
53 --F--T--E--D--P--G--P--D--E--A--P--R--M-- I--E--A--A--P--P--V
73

**YY**M*H**RSR*S*Y*****R**H***YVS***HMYH*YY**R**N**
361 GCCTCTGCTCCAGCTCTACACCGGCGC CTGCTCAGC CCCTCTGCTCCCT 420
219 GGCCCTGACACAGCAGCTCCTACACCGGCGCCCTGACACCGCCCTCCTGGCCCT 278
73 --A--P--A--P--A--A--P--T--P--A--A--P--A--P--A--P--S--W--P--L
93

***D*****W*****BY**DNY*****VD*****RR**B
421 GTCTCTGCTCCAGCTCTACACCGGCGC CTGCTCAGC CCCTCTGCTCCCT 480
279 GTCATCTTCTGCTCCCTTCCAGAAACCTACAGGGCAGCTACGGTTTCGGTCTGGGCTT 338
93 --S--S--S--V--P--S--Q--K--T--Y--Q--G--S--Y--G--F--R--L--G--F
113
    
```

Figura 20: Variante identificada.

- o posición del nucleótido en el cDNA: 140
- o tipo de mutación: : delección de una C (del).
- o Posición del primer aminoácido que cambia y tipo: 47 P (Pro)
- o Consecuencia: cambio en el marco de lectura: frameship (fs)

**Anotación: c.140del (p.Pro47fs).**

*En ClinVar esta variante está clasificada como patológica. **NM\_000546.6(TP53):c.140del (p.Pro47fs).***

*Podemos seguir explorando la variante en ClinVar y vemos que se trata de una mutación que crea un codón de stop prematuro en la posición del aminoácido 76 y que se ha relacionado con el síndrome de Li-Fraumeni. Si nos metemos en Orphanet podemos ver en que consiste la enfermedad: <https://www.orpha.net/consor/cgi-bin/index.php?lng=ES>.*

*En este caso ya no es necesario seguir investigando puesto que la variante es claramente patológica.*

**Repita el proceso con el resto de secuencias disponibles en el repositorio para practicar de forma autónoma y si tienes alguna duda utiliza el FORO.**