

# **Bioinformática y Biología: Investigación sobre el desarrollo de nuevos ARNm como posibles marcadores de la LLC y su relación con el estrés oxidativo**

**Miguel Ángel Romero Fernández-Bravo**  
Máster Universitario en Bioinformática y Bioestadística  
Biología Molecular y Estructural

**Nombre de la Consultora:** Laia Rosich Moya

24 de mayo de 2017

*A mi familia*  
*A mis compañeros: Alicia y Ángel*

*“Todo lo que se puede saber, todavía no se conoce.”*



Esta obra está sujeta a una licencia de  
Reconocimiento-NoComercial-SinObraDerivada  
[3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	<i>Bioinformática y Biología: Investigación sobre el desarrollo de nuevos ARNm como posibles marcadores de la LLC y su relación con el estrés oxidativo</i>
<b>Nombre del autor:</b>	<i>Miguel Ángel Romero Fernández-Bravo</i>
<b>Nombre del consultor/a:</b>	<i>Laia Rosich Moya</i>
<b>Nombre del PRA:</b>	Antoni Pérez Navarro y David Merino Arranz
<b>Fecha de entrega (mm/aaaa):</b>	Mayo/2017
<b>Titulación:</b>	<i>Máster Universitario en Bioinformática y Bioestadística</i>
<b>Área del Trabajo Final:</b>	<i>Biología Molecular y Estructural</i>
<b>Idioma del trabajo:</b>	<i>Español - Castellano</i>
<b>Palabras clave</b>	<i>LLC, Estrés Oxidativo, VIH-1</i>
<p><b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i></p>	
<p>La aplicación de herramientas bioinformáticas para el estudio del cáncer ayuda a entender sus orígenes genéticos. Este estudio trata de la leucemia linfocítica crónica (LLC), una neoplasia hematológica de los linfocitos B.</p> <p>BIRC3 y POT1 son dos genes mutados en pacientes con LLC. Este trabajo se centra en la búsqueda de homólogos a estos genes como posibles marcadores de la LLC. Una búsqueda de los genes asociados con la LLC, devolvió secuencias de VIH-1 en genes asociados al cáncer: BIRC3, POT1, PT53 y OXSR1. El estrés oxidativo puede ser un factor subyacente, ya que está asociado con una serie de neoplasias malignas, incluyendo el gen OXSR1 en pacientes con LLC</p>	
<p><b>Abstract (in English, 250 words or less):</b></p>	
<p>The application of bioinformatics tools to study cancer aids the understanding of their genetic origins. This study involved chronic lymphocytic leukaemia (CLL), a haematological malignancy of B-lymphocytes.</p> <p>BIRC3 and POT1 are two genes mutated in patients with CLL. This paper focuses on the search for homologues to these genes as possible markers of CLL. A search of genes associated with CLL returned HIV-1 sequences in cancer associated genes: BIRC3, POT1, PT53 and OXSR1. Oxidative stress may be an underlying factor since it is associated with a range of malignancies, including OXSR1 gene in patients with CLL.</p>	

# Índice

<b>Capítulo 1. Introducción y Planificación del Trabajo Fin de Máster</b> .....	8
1. Introducción.....	8
1.1 Contexto y justificación del Trabajo.....	8
1.2 Objetivos del Trabajo.....	8
1.3 Enfoque y método seguido.....	9
1.4 Planificación del Trabajo.....	9
<b>Capítulo 2: Desarrollo del Trabajo</b> .....	12
1. Definición de la Leucemia Linfática Crónica (LLC).....	12
1.1 Epidemiología.....	13
1.2 Definición del ARNm.....	13
1.3 Formación de Linfocitos B.....	15
1.4 Marcadores y Estudios Genéticos de la LLC.....	16
1.4.1 Definición de gen.....	17
1.5 Genes asociados a la LLC.....	18
2. Gen BIRC3.....	18
2.1 Localización y secuencia del gen BIRC3.....	19
3. Gen POT1.....	19
3.1 Localización y secuencia del gen POT1.....	20
4. Búsqueda de homologías en el genoma humano.....	21
4.1 Homologías y similaridades del gen BIRC3.....	23
4.1.1 XIAP (X-linked inhibitor of apoptosis protein).....	25
4.1.2 GRIA3 (Glutamate Ionotropic Receptor AMPA type subunit 3).....	28
4.1.3 DOCK7 (Dedicator of Cytokines 7).....	30
4.2. Homologías y similaridades del gen POT1.....	34
4.2.1 SCP2 (Sterol Carrier Protein 2).....	34
4.2.2 PPEF1 (Protein Phosphatase with EF-hand Domain 1).....	37
5. Mutaciones genéticas.....	38
5.1 Mutaciones del gen BIRC3.....	40
5.1.1. Proteína 3T6P-A codificada por el gen BIRC3.....	46
5.2. Mutaciones del gen POT1.....	48
5.2.1 Proteína 3KJP-A codificada por el gen POT1.....	53
6. Estrés Oxidativo.....	55
6.1. Gen OXSR1 “Oxidative Stress Responsive 1”.....	56
6.2. Estrés oxidativo, ARNm y la LLC.....	58
7. Retrovirus y la LLC.....	59
7.1 Homologías entre el gen BIRC3 y retrovirus.....	60
7.2. Retrovirus HTLV-I/II y HTLV-III (VIH-1).....	62
7.2.1 Homologías entre VIH-1 y genes BIRC3 y POT1.....	64
8. LLC y el gen TP53.....	66
8.1. Homología entre el gen TP53 y los genes BIRC3 y POT1.....	68
8.2. Gen TP53 y VIH-1.....	70
<b>Capítulo 3: Conclusiones</b> .....	73
1. Conclusión final.....	73
1.1 Discusión.....	74
2. Reflexión.....	74
3. Análisis crítico.....	74
4. Líneas de trabajo futuras.....	75

<b>Glosario</b> .....	76
<b>Bibliografía</b> .....	77
<b>ANEXO I</b> .....	80
Homología entre el gen OXSR1 y VIH-1 .....	80
<b>ANEXO II</b> .....	83
Homología entre el gen ORF2 y VIH-1 .....	83

## Lista de figuras

Ilustración 1 Tabla de Tareas Planificación Temporal .....	10
Ilustración 2 Diagrama de Gantt .....	11
Ilustración 3: Linfocito B visto al microscopio.....	12
Ilustración 4: Linfocito B afectado por LLC.....	12
Ilustración 5: Transcripción y Traslación Fuente: guiametabolica.org .....	15
Ilustración 6: Diversificación de la Célula Madre .....	15
Ilustración 7: Desarrollo de linfocitos B hasta células memoria y plasmáticas. ....	16
Ilustración 8: Localización de un gen en un cromosoma.....	17
Ilustración 9: Localización del gen BIRC2 en el cromosoma 11.....	19
Ilustración 10: Las mutaciones en POT1.....	20
Ilustración 11: Localización del gen POT1 en el cromosoma 7.....	20
Ilustración 12: Alineamiento correspondiente al resultado TBLASTN del gen BIRC3 c...24	
Ilustración 13: Localización del gen XIAP en el cromosoma X.....	26
Ilustración 14: Resultado de BLASTN entre BIRC3 y XIAP .....	27
Ilustración 15: Localización del gen GRIA3 en el cromosoma X .....	28
Ilustración 16: Resultado de BLASTN entre gen BIRC3 y GRIA3 .....	29
Ilustración 17: Localización de gen DOCK7 en el cromosoma 1.....	30
Ilustración 18: Resultado de BLASTN del gen BIRC3 y DOCK7.....	31
Ilustración 19: Resultado de BLASTN entre los genes GRIA3 y DOCK7 .....	33
Ilustración 20: Resultado de BLASTN entre los genes: BIRC3, XIAP, GRIA3 y DOCK7.....	34
Ilustración 21: Localización del gen SCP2 en el cromosoma 1 .....	35
Ilustración 22: Resultado de BLASTN entre el gen POT1 y SCP2 .....	35
Ilustración 23: Localización del gen PPEF1 en el cromosoma X .....	37
Ilustración 24: Resultado BLASTN entre el gen POT1 y PPEF1 .....	38
Ilustración 25: Ejemplo de mutación de una cadena de ADN por rayos UV. ....	39
Ilustración 26: Tabla de distribución de cánceres asociados a BIRC3 .....	40
Ilustración 27: Tabla de los pacientes que presentan mutación del gen BIRC3.....	41
Ilustración 28: La tabla informativa sobre mutación y tipo de mutación .....	42
Ilustración 29: Información sobre mutaciones dentro de una secuencia de ADN .....	44
Ilustración 30: Información detallada sobre el tipo de mutación .....	45
Ilustración 31: Ejemplo de búsqueda de mutaciones en una secuencia.....	45
Ilustración 32: Ejemplo de delección de dos nucleótidos en el gen BIRC3 .....	46
Ilustración 33: Proteína 3T6P .....	47
Ilustración 34: Imagen de una proteína y su secuencia correcta .....	48
Ilustración 35: Búsqueda en el portal ICGC.....	49
Ilustración 36: Gráfico de de cánceres afectados por la mutación del gen POT1 .....	49
Ilustración 37: Tabla de los pacientes que presentan mutación del gen POT1.....	50
Ilustración 38: La tabla informativa sobre el número de mutación y tipo de mutación	51
Ilustración 39: Resultado que muestra información sobre el tipo de mutación y la región de ADN afectada .....	52
Ilustración 40: Proteína 3KJP-A .....	53
Ilustración 41: Proteína 3KJP consenso.....	54
Ilustración 42: Proteína 3KJP mutada y secuencia de aminoácidos.....	54
Ilustración 43: A la izquierda, la proteína mutada 3KJP. ....	55

Ilustración 44: Localización del gen OXSR1 en el cromosoma 3.....	56
Ilustración 45: La tabla informativa sobre el número de mutación y tipo de mutación	57
Ilustración 46: Resultado sobre el tipo de mutación y la región de ADN afectada .....	57
Ilustración 47: Resultado de la búsqueda BLASTN del gen BIRC3 y retrovirus HERV .....	60
Ilustración 48: Resultado de la búsqueda BLASTN del gen POT1 y retrovirus HERV .....	61
Ilustración 49: Resultado de la búsqueda con BLASTN del virus HTLV-I y HIV-1 .....	63
Ilustración 50: Resultado de la búsqueda con BLASTN del virus HTLV-II y HIV-1 .....	63
Ilustración 51: Resultado de la búsqueda con BLASTN del gen BIRC3 y HIV-1 .....	64
Ilustración 52: Resultado de la búsqueda con BLASTN del gen POT1 y HIV-1 .....	65
Ilustración 53: Localización del gen TP53 en el cromosoma 7 .....	66
Ilustración 54: Tabla de pacientes con mutación en el gen TP53 .....	67
Ilustración 55: La tabla informativa sobre el número de mutación del gen TP53.....	68
Ilustración 56: Resultado de BLASTN del gen TP53 y POT1.....	69
Ilustración 57: Resultado de BLASTN del gen TP53 y BIRC3 .....	70
Ilustración 58: Resultado de la búsqueda en BLASTN del gen TP53 y VIH-1 .....	71

# Capítulo 1. Introducción y Planificación del Trabajo Fin de Máster

## 1. Introducción

### 1.1 Contexto y justificación del Trabajo

La razón por la que he elegido el estudio de la leucemia linfática crónica (LLC), dentro del área de las aplicaciones biomédicas de la secuenciación del genoma, es porque es importante ampliar la búsqueda de posibles biomarcadores y posibles nuevos ARNm que puedan favorecer la detección de la LLC, y entender cómo influye el estrés oxidativo en el desarrollo de nuevos cánceres, en concreto, de la LLC.

La LLC es una enfermedad que suele presentarse en personas de edad comprendida entre los 50 y 70 años, ya que es una enfermedad silenciosa. La detección y diagnóstico temprano de esta enfermedad ayudaría al enfermo a un tratamiento precoz, y así obtener un mayor porcentaje de curación.

Este trabajo, ha tenido como base de partida dos genes relacionados con la LLC: BIRC3 y POT1. Una vez obtenidas las secuencias de ambos genes se han buscado genes homólogos en el cromosoma humano, con el objetivo de buscar posibles nuevos marcadores y ayudar a la predicción de la LLC. Como resultado, se han obtenidos genes homólogos y se han estudiado las mutaciones genéticas, de los genes de estudio, en los pacientes con LLC.

Como consecuencia a la hora de realizar búsquedas en las bases de datos genéticas, se han encontrado homologías con retrovirus, tales como: Human Endogenous Retrovirus (HERV) y Virus de la Inmunodeficiencia Humana (VIH-1).

### 1.2 Objetivos del Trabajo

Los objetivos del trabajo se muestran a continuación en el siguiente listado:

- Definición de conceptos de ARNm.
- Marcadores y estudios genéticos.
- Estudio de los genes mutados como BIRC3 y POT1
- Búsqueda de homologías en el genoma humano.
- Predicción y búsqueda de posibles mutaciones asociables a la LLC
- Estructura molecular de las proteínas codificadas por BIRC3 y POT1
- Simulación molecular.

- Definición de estrés oxidativo.
- Causas o posibles causas del estrés oxidativo.
- Relación entre la oxidación celular y el cáncer.
- Estrés oxidativo y alteración del ARNm.
- Identificación de genes asociados con el estrés oxidativo
- Retrovirus: HERV y VIH
- Homología de los retrovirus con genes asociados a la LLC
- Gen TP53 y la LLC. Homologías con los genes BIRC3 y POT1

### 1.3 Enfoque y método seguido

El trabajo final, se ha enfocado al estudio de la LLC y homólogos de genes asociados a la LLC, es decir, la predicción genética como método de diagnóstico. Al mismo tiempo se ha enfocado el estudio de las mutaciones y cómo estas afectan a la codificación proteica.

Para ello se ha seguido el método de estudio utilizando herramientas bioinformáticas como BLAST (Basic Local Alignment Search Tools), en modo *on line* a través de NCBI (National Center for Biotechnology Information) y utilizando las bases de datos alojadas en el servidor.

Como producto, se han obtenido genes homólogos en base a genes conocidos asociados a la LLC.

Se ha utilizado esta estrategia de búsqueda y no otras herramientas que ofrecen resultados de alineamientos, porque el resultado de la búsqueda en la herramienta *on line* BLAST en NCIB, permite identificar y obtener información del gen que ha sido devuelto como homólogo, de una manera más rápida.

### 1.4 Planificación del Trabajo

Durante el desarrollo del trabajo, hubo desviaciones causadas por factores externos que se contemplaron en el plan de contingencia, aun así, hubo que reestructurar el trabajo y los temas a desarrollar.

Para desarrollar el trabajo se han empleado las siguientes herramientas:

- Ordenador Apple MacBook Pro Intel Core i7 a 2.2 GHz. con 16 Gb de Memoria RAM y Sistema Operativo MacOS X 10.12.4.
- Software libre GanttProject.
- Editor de Textos Word.
- BLAST Basic Local Alignment Search Tool en versión online, cuando sea posible, en: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

- Bases de datos genómicas como: NCBI, EMBL, Uniprot, PDB, Swiss-Prot, SCOP, ENSEMBL.
- Base de datos de genéticas: OASIS, PDB, Genecards, ICGC, NCBI, GHR, Cancer3D.
- Software de simulación de proteínas y moléculas, visor de estructuras químicas y biomoléculas: Chimera.

A continuación, se muestra en una tabla la planificación temporal:

## Tarea

Nombre	Fecha de inicio	Fecha de fin
01 Trabajo Fin de Master	1/03/17	1/03/17
01.01 Planificación de Trabajo TFM (PEC1)	1/03/17	14/03/17
01.02 Entrega Plan de Trabajo TFM (PEC1)	15/03/17	15/03/17
02 Definición de la LLC y nuevos ARNm	15/03/17	17/03/17
02.01 Causas de la LLC y epidemiología	18/03/17	18/03/17
02.01.01 Homologías entre los genes asociados	18/03/17	20/03/17
02.01.02 Búsqueda de homólogos dentro de la secuenciación del genoma humano	21/03/17	8/04/17
02.02 Concepto de ARNm	9/04/17	9/04/17
02.03 Revisión Final Documentación - Borrador	9/04/17	25/04/17
02.04 Entrega Finalizada PEC 2	26/04/17	26/04/17
03 Mutaciones	26/04/17	28/04/17
03.01 Mutaciones asociadas a BIRC3 y POT1	29/04/17	3/05/17
03.02 Simulación Molecular	4/05/17	6/05/17
03.03 Estrés Oxidativo	7/05/17	7/05/17
03.03.01 Definición del estrés oxidativo	7/05/17	9/05/17
03.03.02 Causas del estrés oxidativo	10/05/17	10/05/17
03.04 Gen OXSR1	11/05/17	11/05/17
03.04.01 Estrés oxidativo y alteración del ARNm. Identificación de posibles secuencias	12/05/17	12/05/17
03.05 Gen OXSR1 la LLC	13/05/17	13/05/17
03.06 Retrovirus y la LLC	14/05/17	14/05/17
03.06.01 TP53 y la LLC	15/05/17	15/05/17
03.06.02 Homología entre los genes TP53, BIRC3 y POT1	16/05/17	16/05/17
03.07 Revisión de la documentación - Borrador	17/05/17	17/05/17
03.08 Entrega de la Revisión PEC 3	18/05/17	18/05/17
04 Elaboración Memoria y Presentación TFM	18/05/17	18/05/17
04.01 Elaboración Memoria	19/05/17	19/05/17
04.02 Elaboración Presentación TFC	20/05/17	20/05/17
04.03 Revisión Presentación TFM	21/05/17	22/05/17
04.04 Entrega Memoria y TFM Entrega Final	23/05/17	24/05/17

*Ilustración 1 Tabla de Tareas Planificación Temporal*

También se observa la planificación de forma más resumida en el siguiente diagrama de Gantt:

## Diagrama de Gantt

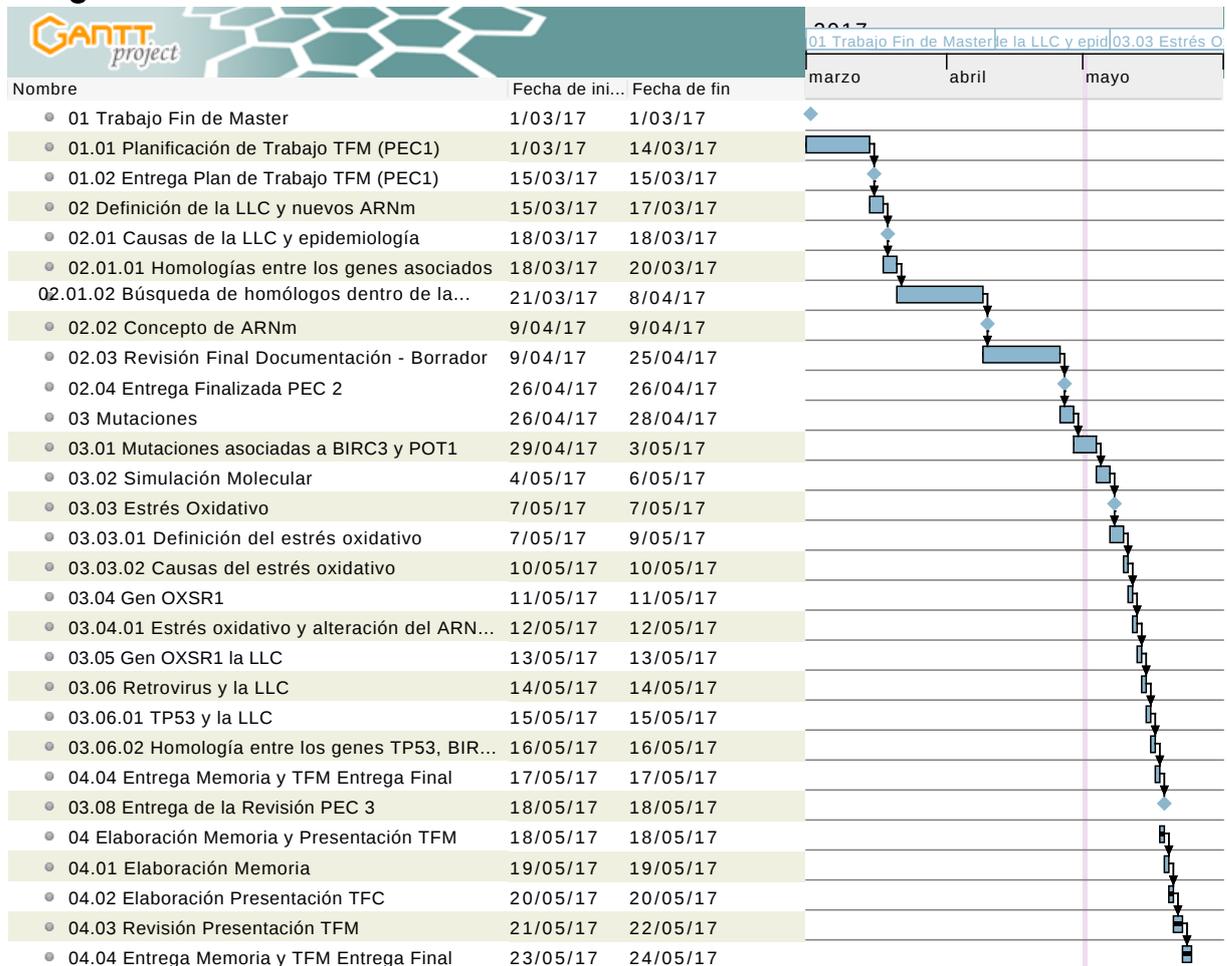


Ilustración 2 Diagrama de Gantt

## Capítulo 2: Desarrollo del Trabajo

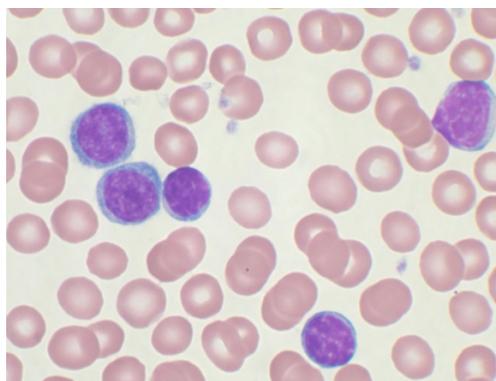
### 1. Definición de la Leucemia Linfática Crónica (LLC)

La leucemia es un tipo de cáncer que afecta a la sangre. En concreto la leucemia linfática crónica (también es conocida como leucemia linfocítica crónica, LLC), es un tipo de leucemia que afecta a los glóbulos blancos (específicamente a los linfocitos B), es decir, células se encargan del mecanismo de defensa del cuerpo frente a agentes externos: virus, bacterias, parásitos, hongos, etc. Este cáncer se origina debido a que los linfocitos que se generan en la médula ósea empiezan a crecer de una forma descontrolada y pasan al torrente sanguíneo donde se acumulan lentamente. Las personas afectadas pueden no notar los síntomas durante mucho tiempo, incluso años. Con el tiempo, este tipo de células pueden propagarse a otras partes del cuerpo en las que se incluyen los ganglios linfáticos, el hígado y el bazo.

A continuación, se observan dos figuras en las que se muestran un linfocito normal (Ilustración 3) y un linfocito de LLC (Ilustración 4):



*Ilustración 3: Linfocito B visto al microscopio, en el que se observa una cobertura del núcleo normal y bien definido.*



*Ilustración 4: Linfocito B afectado por LLC. Si se observa bien, podemos ver un núcleo con líneas parecido al caparazón de una tortuga.*

Cuando una célula tipo linfocito no madura normalmente, se divide y se multiplica de forma anormal y no muere cuando debería de hacerlo, por lo que se convierte en una célula leucémica que se acumula en la médula ósea, desplazando a las células normales. Esto ocurre en un determinado momento provocando un aumento en el número de glóbulos blancos en el torrente sanguíneo, lo que hace que este tipo de cáncer sea diferente a otros localizados, que empiezan a desarrollarse en órganos como: pulmones, páncreas, seno, etc., y que luego se

propagan a la médula ósea. Este tipo de cáncer empieza desde la médula ósea y se expande a otros órganos.<sup>(1)</sup>

En el caso de la LLC, se observa una acumulación de linfocitos B neoplásicos que se hallan en la fase inicial del ciclo celular que expresan CD5, CD19, CD20 y CD23<sup>1</sup>. La causa de la LLC sigue siendo desconocida, pero diferentes condiciones se han identificado como factores de riesgo de padecerla, por ejemplo:<sup>(2)</sup>

- Antecedentes familiares de cáncer y neoplasias hematológicas.
- Inmunodeficiencias.
- Exposición al agente naranja (herbicida).
- Edad avanzada.
- Alteraciones cromosómicas específicas.

## 1.1 Epidemiología

Se considera la LLC como la neoplasia hematológica más frecuente en los países occidentales, elevándose su incidencia conforme aumenta la edad. Es poco frecuente antes de los 50 años con una incidencia de 5/100.000 habitantes/año y representa menos del 20% de todos los diagnósticos, aunque a partir de los 70 años se observa un aumento de forma importante de 30/100.000 habitantes/año. Podemos observar que la edad media de diagnóstico es de 68-70 años y es más frecuente en varones que en mujeres (2:1) y entre el 40 – 50% de los pacientes son mayores de los 70 años.<sup>(3)</sup>

## 1.2 Definición del ARNm

El ARN, es el ácido ribonucleico presente en las células tanto eucariotas como procariotas y también en algunos tipos de virus. El ARN está formado por una cadena simple de bases nitrogenadas: la Adenina (A), la Guanina (G), Citosina (C) y el Uracilo (U). La diferencia de bases que existe entre el ADN y el ARN, es que el ADN tiene una base nitrogenada llamada Timina (T). El ARN es una cadena lineal monocatenaria, es decir, es una cadena lineal simple de ácidos ribonucleicos. El ARN desempeña diversas funciones ya que el ADN se encuentra en el núcleo celular y no puede salir.<sup>(4)</sup>

Dentro de la célula se encuentra el núcleo celular que se encarga de proteger la información genética y es donde se encuentran los cromosomas formados por cadenas plegadas de ADN. Al no poder salir de la célula, se necesita un

---

<sup>1</sup>Las siglas CD corresponden a "Cluster of Differentiation" o cúmulo de diferenciación, que son moléculas marcadoras de la superficie celular que son reconocidas por ciertos anticuerpos y son usadas para la identificación del tipo de célula, estadio de diferenciación celular y la actividad de la misma. Se caracterizan por anticuerpos monoclonales y así se permite caracterizar los tipos de leucocitos. Por ejemplo, en el caso de CD5 se caracteriza por tener un peso molecular de 55 KDa y su función biológica es mediar en la transducción de la señal<sup>(2)</sup>.

mecanismo que transfiera la información genética, en este caso, se utiliza el ARN que es la cadena complementaria del ADN.

El ARN mensajero (ARNm) es el producto obtenido de la transcripción de una cadena de ADN de un cromosoma que codifica a un gen. La cadena de ARNm es procesada por el ribosoma de la célula quien permitirá convertir el ARNm en un patrón de aminoácidos, estos aminoácidos serán los que formen la proteína.

El proceso de formación del ARNm y su posterior traducción es algo que parece sencillo a simple vista, pero es un proceso bastante complejo, ya que para generar la cadena final del ARNm primero se ha de procesar una cadena pre-ARN que viene directamente de la transcripción del ADN del núcleo celular. Es importante comprender bien este proceso, ya que es el proceso de transferencia de información genética del ADN y puede haber fallos que desencadenen proteínas defectuosas y, en nuestro caso de estudio, pueda ser probable causa de LLC por una posible mutación de la cadena de ARNm debido a factores externos.

Para obtener una proteína, se han de seguir unas fases para que no haya fallos a la hora de la síntesis. La cadena de ADN que codifica el gen consta de señales de inicio y de fin de cadena, así como de intrones y de exones.

En el proceso de formación del ARNm, se obtiene una cadena primaria que se llama Pre-ARNm mediante el proceso de transcripción que consiste en la eliminación de los intrones de la cadena de ADN inicial, después se realiza el proceso de splicing (corte y empalme) que unirá los exones del ADN, que son los que codificarán la proteína<sup>(5)</sup>. Por último, el proceso final será el de traslación o traducción que consiste en que el ribosoma lee la cadena final del ARNm y obtiene la secuencia de aminoácidos que codifica la proteína.

En la siguiente figura, se observa cómo se realiza el proceso de maduración del ARNm para obtener la proteína final desde la cadena inicial de ADN:

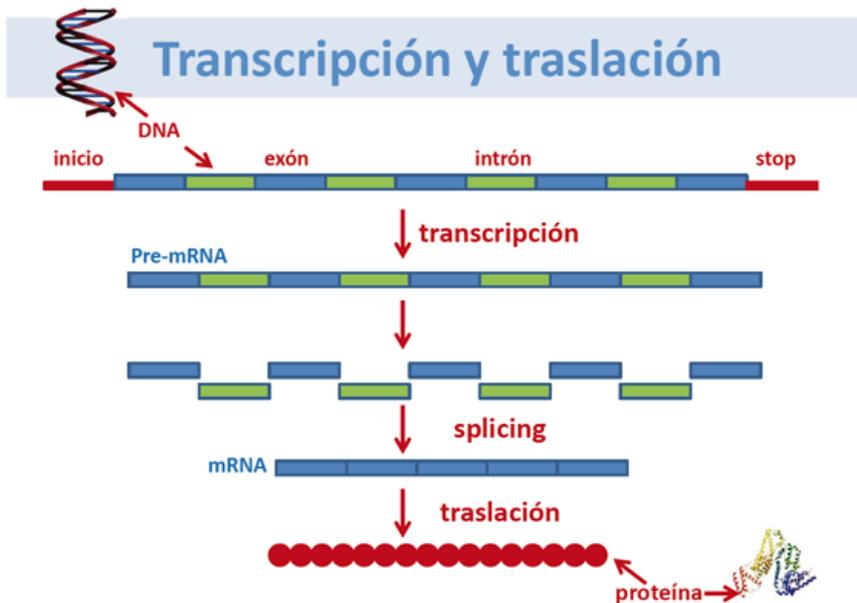


Ilustración 5: Transcripción y Traducción Fuente: [guiametabolica.org](http://guiametabolica.org)

### 1.3 Formación de Linfocitos B

Es necesario conocer la formación de linfocitos B (o células B) para entender el proceso que causa la leucemia. Los linfocitos B proceden de la médula ósea (Bone Marrow, de ahí proviene el nombre de linfocitos B) y es en la médula ósea donde se encuentran células madre sanguíneas. Éstas se pueden convertir en células madre mieloide (glóbulos rojos, plaquetas, granulocitos) o células madre linfoide, que pueden convertirse en linfocitos B como podemos ver en la siguiente Ilustración 6:

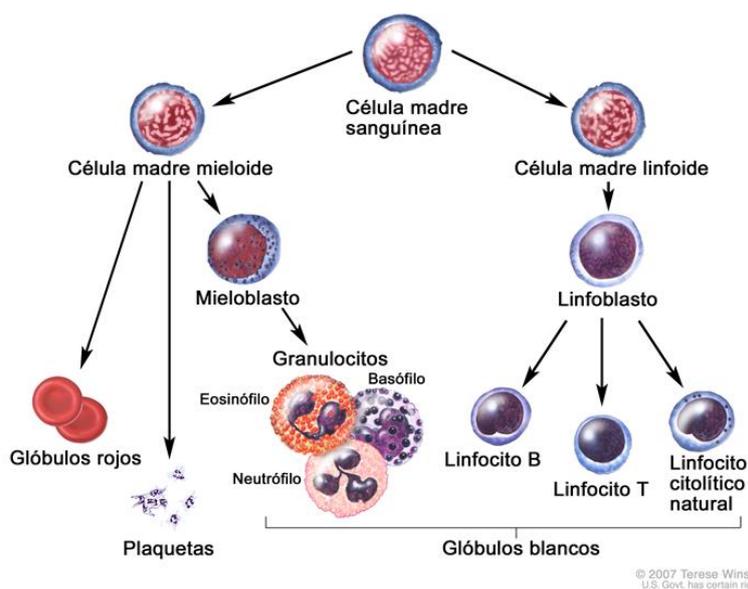


Ilustración 6: Diversificación de la Célula Madre

Los linfocitos B, se forman a partir de progenitores linfoides derivados de las células hematopoyéticas primordiales (HSC) mediante un proceso que se llama linfopoyesis B. Las células hematopoyéticas primordiales (también son llamadas células madre hemapoyéticas), son un grupo de células inmaduras que se pueden transformar en todos los tipos de células sanguíneas, como linfocitos, plaquetas y glóbulos rojos. Este tipo de células se encuentran en la sangre periférica y en la médula ósea. La maduración de los linfocitos B se comenta brevemente para entender el proceso y consta de tres fases (Ilustración 7):

- a) Células primitivas B.
- b) Formación de células pre-B.
- c) En la última fase se forman los linfocitos B maduros que son capaces de expresar inmunoglobulinas del tipo IgM e IgD <sup>(6)</sup>.

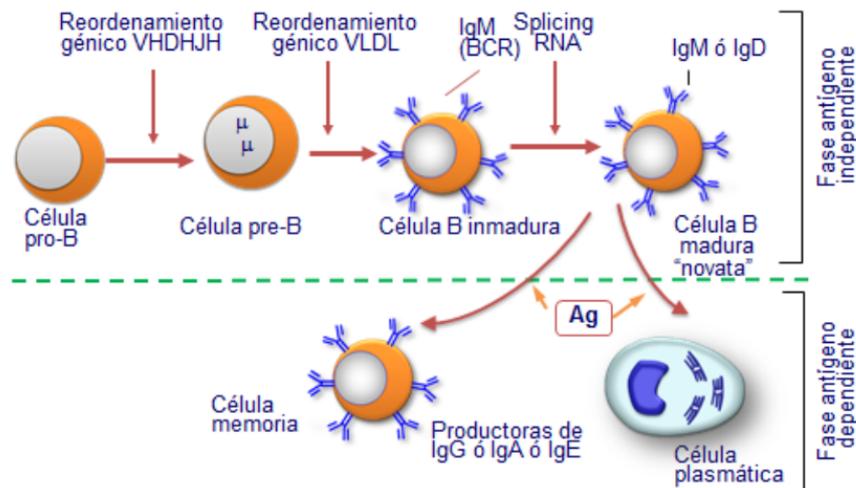


Ilustración 7: Desarrollo de linfocitos B hasta células memoria y plasmáticas.

## 1.4 Marcadores y Estudios Genéticos de la LLC

Los proyectos actuales de la secuenciación del genoma humano han ayudado a la investigación de enfermedades relacionadas con el cáncer, como el caso de la LLC. La secuenciación del genoma de la LLC<sup>(7)</sup> ha sido un hito en la investigación de las posibles causas de la LLC en la población como anteriormente se ha comentado. Aunque las mutaciones genéticas no son la única causa para desarrollar la LLC, sí es cierto que su estudio puede ayudar a la prevención y al tratamiento de la misma.

Actualmente y gracias a los grupos de investigación, existen bancos de datos genómicos con información de los genes implicados en el desarrollo de diferentes tipos de cáncer, en los que se incluyen la LLC, y también se dispone de secuencias genéticas que ayudan al estudio de las mutaciones.

Usando esta información, se pueden crear unos marcadores que podrán ser utilizados para el diagnóstico y el entendimiento del origen de la LLC, consiguiendo un tratamiento precoz de la enfermedad.

#### 1.4.1 Definición de gen

Una cadena de ARNm contiene la información de un gen, de modo que podemos entender que un gen es una secuencia de ADN que contiene información para codificar a una proteína. Los genes están localizados en cromosomas, en concreto, el ser humano tiene 23 pares de cromosomas, cada cromosoma está formado por un centrómero, que es la región estrecha del cromosoma y separa a este en dos partes: la parte corta que se anota como 'p' y la parte larga que se anota como 'q'. También consta de bandas y sub-bandas que permiten localizar dónde se encuentra el gen.

Se muestra un ejemplo:

- En la siguiente imagen, se observa la localización de un gen llamado CFTR y la anotación sería: 7q31.2, es decir, el gen se encuentra en el cromosoma 7, en el brazo q (largo), en la región 3, en la banda 1 y en la sub-banda 2. <sup>(8)</sup>

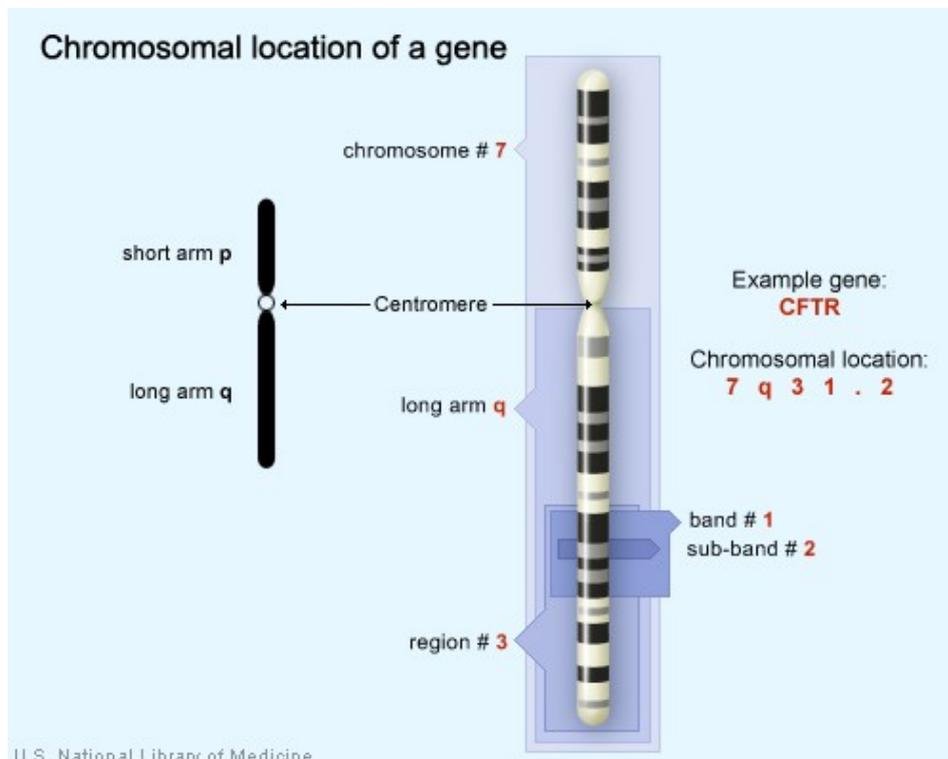


Ilustración 8: Localización de un gen en un cromosoma

## 1.5 Genes asociados a la LLC

Hoy en día se conocen algunos genes asociados a la LLC como NOTCH1, XPO1, MYD88 o KLHL6, aunque este estudio se centra en los genes BIRC3 y POT1.

También, hay alteraciones citogenéticas que pueden ser causantes de una LLC, como: deleción<sup>2</sup> en el cromosoma 13q14, deleción en el cromosoma 11q22-23, trisomía del cromosoma 12 o mutación del gen TP53<sup>(7)</sup>.

A continuación, se estudiará con más detenimiento los genes BIRC3 y POT1, cuyas alteraciones pueden estar asociadas a la LLC.

## 2. Gen BIRC3

El gen BIRC3 (Baculoviral IAP Repeat Containing 3), es conocido por su función como inhibidor de la apoptosis (IAP). La apoptosis consiste en la muerte celular programada con la finalidad de controlar el desarrollo y el crecimiento celular.

La apoptosis puede ser desencadenada por señales celulares que se controlan genéticamente. El gen BIRC3 pertenece a este grupo de genes y la proteína codificada, aparte de inhibir la apoptosis, también modula la señalización inflamatoria y la inmunidad a través de la señalización de la quinasa mitogénica, que es una enzima que opera como transductor de señales.

El gen BIRC3 también está implicado en la proliferación celular, de la invasión de células y de la metástasis<sup>(9)</sup>. En consecuencia, si se produce una mutación o un mal funcionamiento de este gen, se puede llegar a un mal funcionamiento del factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas (NFKB), ya que la mutación del gen BIRC3 hace que el mecanismo NFKB esté siempre activo<sup>(10)</sup>.

El mecanismo NFKB es un complejo proteico que controla la transcripción del ADN y está implicado en la respuesta celular frente a estímulos como el estrés oxidativo, citoquinas, radiación ultravioleta, antígenos bacterianos o virales<sup>(11)</sup>.

---

<sup>2</sup>La deleción es la pérdida del material genético de un cromosoma, también puede entenderse como mutación y puede ir desde perder pequeñas áreas hasta grandes regiones.

## 2.1 Localización y secuencia del gen BIRC3

BIRC3 es un gen que se encuentra en el cromosoma 11 en la posición 22.2, su anotación es: 11q22.2. Su secuencia corresponde con los pares de bases de 102.317.450 a 102.339.403 en el cromosoma 11.

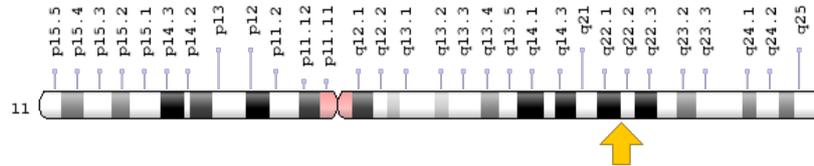


Ilustración 9: Localización del gen BIRC2 en el cromosoma 11

La anotación de este gen la podemos encontrar anotada en forma WGS (Whole Genome Shotgun) en GenBank, con el número de acceso NC\_018922.2<sup>(12)</sup>. También se encuentra la misma secuencia anotada en la región del cromosoma 11, según la actual secuenciación del proyecto Genoma Humano GRCh38.p7, con anotación en GenBank: NC\_000011.10<sup>(13)</sup>. Ambas secuencias tienen 21.956 pares de bases (bp).

## 3. Gen POT1

El gen POT1 (Protección de Telómeros 1), codifica la proteína nuclear implicada en el mantenimiento de los telómeros que pertenece a un complejo multiproteico que se llama telosoma, que se une a las repeticiones TTAGGG de los telómeros y así regula la longitud y protege los extremos del cromosoma de una recombinación ilegítima, también de una inestabilidad cromosómica catastrófica o de segregación cromosómica anormal<sup>(14)</sup>.

Las recientes investigaciones que se han centrado en la LLC, han dado como resultado que mutaciones en el gen POT1 se encuentran en pacientes con este tipo de leucemia. Debido a que cada cromosoma tiene telómeros en su extremo, que es como una cubierta formada de proteínas, el gen POT1 es el que se encarga de anclarlo en su sitio uniéndolo al ADN telomérico, por lo que las mutaciones encontradas en POT1 impiden que el gen cumpla con su función, llevando así a una anomalía en el crecimiento descontrolado de los linfocitos B como podemos ver en la Ilustración 10:<sup>(15)</sup>

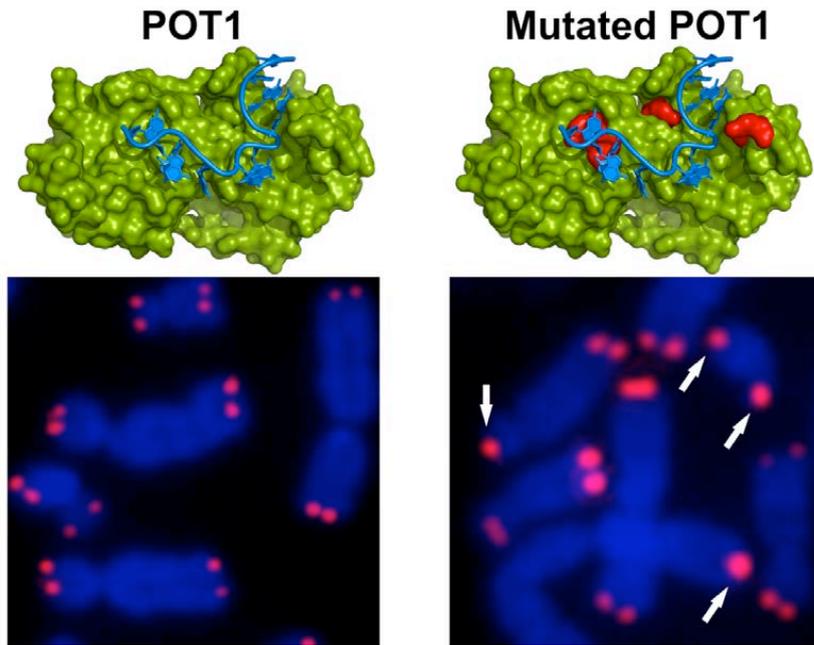


Ilustración 10: Las mutaciones en *POT1* generan alteraciones cromosómicas. En la parte superior izquierda se representa la estructura de *POT1* (verde) unida a ADN telomérico (azul). En condiciones normales (debajo), durante la metafase los cromosomas (en azul) presentan dos telómeros (en rojo) en cada extremo, uno por cromátida. En la parte superior derecha se encuentra la misma estructura, con las zonas mutadas, todas ellas cercanas al ADN telomérico, señaladas en rojo. Debajo se muestran imágenes de cromosomas durante la metafase en células con mutaciones en *POT1*. En este caso, se puede observar que muchos cromosomas tienen un solo telómero en su extremo (flechas blancas). Fuente: <https://www.cnio.es/es/news/docs/maria-blasco-nature-genetics-17mar13-es.pdf>

### 3.1 Localización y secuencia del gen *POT1*

*POT1* es un gen que se encuentra en el cromosoma 7 en la posición q31.33 anotado como: 7q31.33. Su secuencia corresponde a los pares de bases que van desde 124.822.386 hasta 124.929.983, según la última anotación del genoma humano GRCh38.p7.

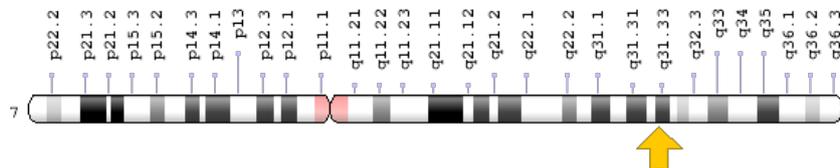


Ilustración 11: Localización del gen *POT1* en el cromosoma 7

La notación del gen *POT1* la podemos encontrar con el número de acceso NG\_029232.1<sup>(16)</sup> en GenBank. También, se encuentra la secuencia completa acotada según la anotación GRCh38.p7 y su número de acceso es: NC\_000007.14<sup>(17)</sup>. Ambas secuencias constan de 107.598 bp.

## 4. Búsqueda de homologías en el genoma humano

Para la búsqueda de genes o secuencias homólogas en el genoma humano, se utilizan herramientas y métodos que se describen a continuación.

Una de las herramientas que se utilizarán para las búsquedas en las bases de datos es BLAST (Basic Local Alignment Search Tool o búsqueda básica de alineamientos locales), que es un software ampliamente investigado y utilizado en bioinformática, y calcula el grado de relación entre los nucleótidos de secuencias de ADN o las secuencias de aminoácidos (AA) de diferentes proteínas.

Estas secuencias comparadas se denominan secuencias de consulta (“*query*”) y sujeto (“*subject*”). Los algoritmos que se utilizarán en esta búsqueda serán BLASTN, BLASTX y TBLASTN. BLASTN, compara dos secuencias de nucleótidos (ADN), BLASTX compara una secuencia consulta en nucleótidos con bases de datos que contienen proteínas (AA), traduciendo los nucleótidos de la consulta en AA y después comparando.

Por último, TBLASTN es un algoritmo que una vez introducida una secuencia consulta en AA, lo compara con las bases de datos que se encuentran en nucleótidos, traduciendo el sujeto en una secuencia AA (proteína) en todos los 6 marcos de lectura y luego compara con la consulta.

TBLASTN ha mejorado la sensibilidad y la importancia biológica en comparación con BLASTN. Brevemente, BLAST genera y compara las subcadenas de la secuencia de consulta con subcadenas de las secuencias sujeto. Las comparaciones que superan una determinada puntuación umbral (“hits”) se ordena de acuerdo a su longitud y significado y se devuelven en forma de formulario<sup>(18)</sup>.

Los resultados se interpretan de la siguiente forma:

1. *Cover*, es el porcentaje de la secuencia de consulta que está alineada con la secuencia sujeto.
2. *Identities*, es el número y porcentaje de coincidencias exactas entre la consulta y el sujeto sobre la longitud de la secuencia.
3. *Similarities*, es el número y porcentaje de coincidencias exactas más sustituciones conservadas para AA. Las sustituciones conservadas se refieren a aminoácidos con propiedades fisicoquímicas similares, por ejemplo, la leucina e isoleucina.
4. *Expecte value E*. El cálculo del valor *E-Value* es un método para evaluar el significado estadístico de la alineación y es necesario para determinar alineaciones biológicamente relevantes y separarlos del ruido aleatorio.

El valor *E-value* es el número de coincidencias con la misma puntuación que se espera que se produzca por casualidad dentro de una base de datos determinada (19) (20). Por ejemplo, un valor *E-value* de  $10^{-6}$  equivale a una coincidencia esperada por azar para cada 1.000.000 entradas en la base de datos. Cuanto más bajo sea el valor de *E-value* o más cerca esté de cero, más significativa será la coincidencia.

Las alineaciones de nucleótidos con valor  $E < 10^{-6}$  e identidad  $\geq 70\%$ , se consideran significativas de homología. Las secuencias de proteínas de más de 100 AA de longitud, con valores de alineación  $E < 10^{-3}$  e identidad  $\geq 25\%$  son también significativas y sugieren similitud. Las proteínas homólogas "probablemente tienen el mismo antepasado, comparten la misma estructura y tienen una función biológica similar"<sup>(20)</sup>. La homología no debe confundirse con similitud o identidad. La homología, aunque basada en estas variables, está presente o ausente. Por ejemplo, es incorrecto afirmar que dos proteínas son homólogas al 90%.

En la búsqueda de homólogos, no sólo se ha limitado al genoma humano, es decir, a la especie *Homo sapiens*, sino que también se ha buscado homólogos en *Homo neanderthalensis*, pero no se ha encontrado ningún resultado positivo. La búsqueda también se ha ampliado a los siguientes organismos: los virus, los hongos, los parásitos y las bacterias. Esta última búsqueda puede hacernos comprender ancestros comunes a los genes BIRC3 o POT1.

Se utilizarán dos herramientas para poder consultar posibles homologías. Una de ellas es BLAST, que está alojada en National Center for Biotechnology Information (NCBI) que contiene un repositorio de bases de datos que son enlazadas desde la aplicación online BLAST. Las bases de datos enlazadas que utilizarán son:

- Human RefSeqGene sequences (RefSeq\_Gene), esta base de datos contiene secuencias del genoma a consultar, devolviendo como respuesta secuencias identificadas con genes.
- RefSeq Representative genomes, es una base de datos que contiene genomas representativos.
- Reference RNA sequences (refseq\_rna), la base de datos refseq\_rna devuelve secuencias RNA, las cuales son importantes para el estudio del mRNA.
- Nucleotide Collection (nr/nt), es la base de datos que contiene multitud de secuencias, es como realizar una consulta en general.
- NCBI Genomes (chromosomes), es una base de datos que devuelve como resultado alineaciones correspondientes a cromosomas.

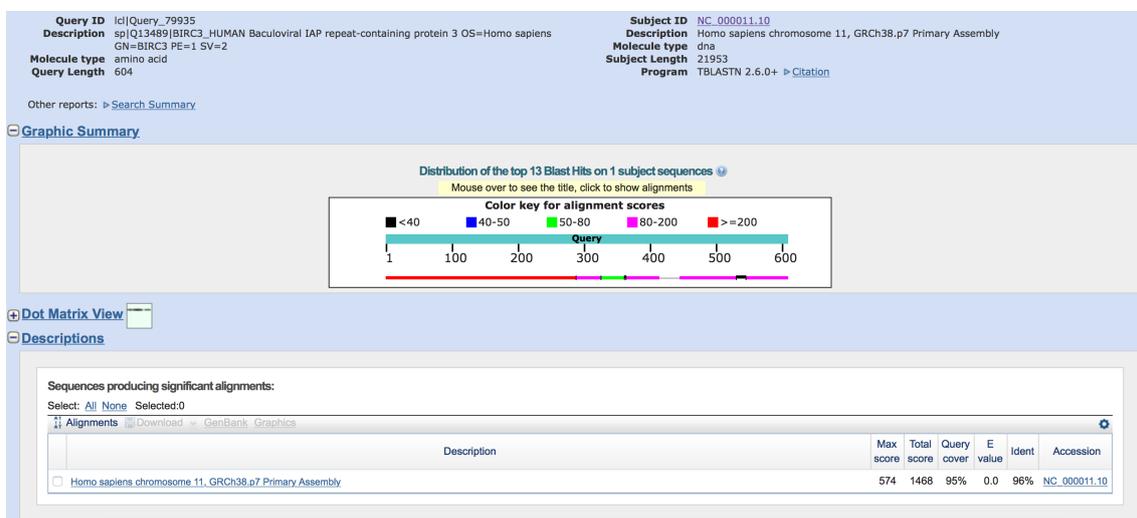
Otra herramienta que se utilizará para la búsqueda es BLAT de la Universidad de Santa Cruz (USCS), California EEUU. BLAT, está diseñado para una búsqueda más rápida (alrededor del 95%) de secuencias, con una mayor similitud de secuencias con longitud de 25 o 40 bases o más.

En cuanto a la búsqueda de proteínas, encuentra longitudes de 20 o más aminoácidos. Aunque BLAT aparentemente se parece a BLAST, porque ambas son herramientas de alineación, están estructuradas de una forma diferente. En las secuencias de ADN, BLAT funciona bajo un índice de un genoma completo que está en memoria, por lo que la base de datos objetivo de BLAT no es un conjunto de secuencias como puede ser las contenidas en GenBank, sino que devuelve un índice derivado del ensamblaje de todo el genoma.<sup>(21)</sup>

#### 4.1 Homologías y similitudes del gen BIRC3

Durante el proceso de búsqueda de información de secuencias homólogas correspondientes a BIRC3 en anotación de nucleótidos (NC\_018922.2), se realizó la búsqueda al mismo tiempo del gen BIRC3 en secuencia de aminoácidos y así ampliar la búsqueda de homólogos. Sorprendentemente, la proteína secuenciada no corresponde a la secuencia de ADN de la región que codifica la proteína anotada en pares de bases, según la última secuenciación del proyecto genoma humano.

A continuación, se muestra el resultado de la búsqueda TBLASTN. En la imagen se observa que se ha introducido la secuencia de la base de datos UniProtKB/Swiss-Prot correspondiente al número de acceso: Q13489.2 y se ha comparado con la secuencia origen NC\_000011.10 en la región 102.317.450 bp hasta 102.339.403 bp, que corresponde al gen BIRC3, obteniendo el siguiente alineamiento:



Download GenBank Graphics Sort by: E value

Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly  
Sequence ID: NC\_000011.10 Length: 135086622 Number of Matches: 13

Range 1: 102324510 to 102325364 GenBank Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
574 bits(1479)	0.0	Compositional matrix adjust.	285/285(100%)	285/285(100%)	0/285(0%)	+3
Query 1		MNIVENSIFLSNLMKSANTFELKYDLSCELYRMSTYSTFPAGVPSERSLARAGFYTTGV				60
Sbjct 102324510		MNIVENSIFLSNLMKSANTFELKYDLSCELYRMSTYSTFPAGVPSERSLARAGFYTTGV				102324689
Query 61		NDKVKFCFCGLMLDNWKRGDSPTEKHKLYPSCRQVQSLNSVNNLEATSQPTFPSSVTNS				120
Sbjct 102324690		NDKVKFCFCGLMLDNWKRGDSPTEKHKLYPSCRQVQSLNSVNNLEATSQPTFPSSVTNS				102324869
Query 121		THSLLPGTENSIFYRGSYSNSPSNPVNSRANQDFSAIMRSSHHCAMNNENARLLTFQTWP				180
Sbjct 102324870		THSLLPGTENSIFYRGSYSNSPSNPVNSRANQDFSAIMRSSHHCAMNNENARLLTFQTWP				102325049
Query 181		LTFLSPTDLAKAGFYIIGPDRVACFACGGKLSNWEPKDNAMSEHLRHFPCPFIEHQLO				240
Sbjct 102325050		LTFLSPTDLAKAGFYIIGPDRVACFACGGKLSNWEPKDNAMSEHLRHFPCPFIEHQLO				102325229
Query 241		DTSRYTVSNLSMQTHAARFKTFNWPSSVLVNPQLASAGFYVVG	285			
Sbjct 102325230		DTSRYTVSNLSMQTHAARFKTFNWPSSVLVNPQLASAGFYVVG		102325364		

Range 2: 102335965 to 102336219 GenBank Graphics Next Match Previous Match First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
171 bits(434)	6e-48	Compositional matrix adjust.	84/85(99%)	85/85(100%)	0/85(0%)	+1
Query 442		NDLLLRKRNRMALFQHLTCVPIPLDLSLLTAGIINEQEHVDVIKQKQTQSLQARELIDTILV				501
Sbjct 102335965		+DLLLRKRNRMALFQHLTCVPIPLDLSLLTAGIINEQEHVDVIKQKQTQSLQARELIDTILV				102336144
Query 502		KGNIATVFRNSLQEAQAVLYEHLF	526			
Sbjct 102336145		KGNIATVFRNSLQEAQAVLYEHLF		102336219		

Range 3: 102336908 to 102337099 GenBank Graphics Next Match Previous Match First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
109 bits(273)	1e-30	Compositional matrix adjust.	64/64(100%)	64/64(100%)	0/64(0%)	+2
Query 541		DLPVEEQLRRLQEERTCKVCMDEKVSIVFIPCGHLVVKDCAPSLRKCPICRSTIKGTVR				600
Sbjct 102336908		DLPVEEQLRRLQEERTCKVCMDEKVSIVFIPCGHLVVKDCAPSLRKCPICRSTIKGTVR				102337087

Ilustración 12: Alineamiento correspondiente al resultado TBLASTN del gen BIRC3 con número de acceso Q13489.2

Se observa observar lo siguiente:

- **Query:** corresponde al gen BIRC3 en secuencia de aminoácidos con una longitud de 604 aa.
- **Subject:** corresponde al gen BIRC3 en secuencia de nucleótidos con una longitud de 21.953 bp. TBLASTN se encarga de traducir *subject* en aminoácidos siguiendo el método anteriormente explicado y después lo compara con *query*.
- Observamos el campo *subject* y la longitud de la secuencia del gen en nucleótidos <sup>(13)</sup>. Si se convierte la longitud en aminoácidos, devuelve como resultado 7.318 aa que es mayor que la secuencia *query* que codifica el gen BIRC3. Esto puede ser debido al proceso de transcripción del ADN al ARNm, por la delección de los intrones. De ahí que el alineamiento no esté ajustado.
- La longitud máxima de la secuencia que devuelve una similaridad del 100% es de 285 aa de la proteína *query* con respecto a *subject*.
- Los valores del alineamiento corresponden a un *E-value* de 0.0, con un *identity* del 96% y un *cover* del 95%, aunque el alineamiento obtiene unos valores muy buenos, para realizar la búsqueda de homólogos se va a optar a usar la secuencia en ADN, no la secuencia en proteína Q13489

ya que se puede perder información a la hora de buscar regiones homólogas y similitudes.

- Cabe destacar que el alineamiento empieza con la señal de inicio del gen, codificado con el aminoácido Met, en la proteína Q13489 (*query*) con respecto a la secuencia en nucleótidos traducidos (*subject*), pero hay saltos en el alineamiento.

Debido a estos motivos, se utilizará la secuencia en nucleótidos para la búsqueda de homólogos. El proceso consistirá en introducir de nuevo la secuencia de ADN con número de acceso: NC\_000011.10 en la región definida por los pares de bases 102.317.450 hasta 102.339.403.

En la búsqueda, se han consultado varias bases de datos y se ha obtenido como resultado el mismo gen BIRC3 con nombres alternativos tales como: MIHC o API2, que se refieren al mismo gen BIRC3 y están localizados en el mismo cromosoma y posición.

Utilizando BLASTN para la búsqueda, se ha hallado homología con los genes que se estudiarán en los siguientes apartados.

#### 4.1.1 XIAP (X-linked inhibitor of apoptosis protein)

El gen XIAP, se encarga de codificar una proteína que se encuentra en las células inmunitarias. Este tipo de gen protege a estas células de una autodestrucción por inhibición de ciertas enzimas llamadas caspasas, las cuales son necesarias para la apoptosis. La mutación del gen XIAP ha sido identificada en algunas personas con la enfermedad linfoproliferativa ligada al cromosoma X (XLP), que es una alteración del sistema inmune y de las células sanguíneas, afecta exclusivamente a los varones los cuales, pueden llegar a desarrollar linfomas. Cuando el gen XIAP está mutado, puede o bien no codificar la proteína o bien no fabricar la cantidad de proteínas necesarias. Esta enfermedad se hereda con un patrón recesivo ligado al cromosoma X.<sup>(22)</sup>

El gen XIAP, está situado en el cromosoma Xq25 y la región del cromosoma X que lo contiene, está comprendida entre los pares de bases 123.859.812 y 123.913.979, según la anotación GRch38p7. El número de acceso en GenBank es: NG\_007264.1 .

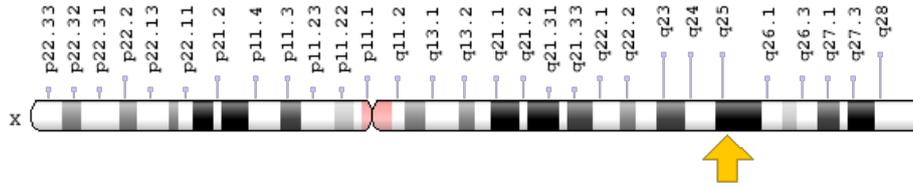


Ilustración 13: Localización del gen XIAP en el cromosoma X

A continuación, se muestra una parte del alineamiento. El alineamiento completo se puede ver aquí: <http://goo.gl/3epQzo> debido a que el resultado es muy grande por la longitud de ambos genes.

**Query ID** NC\_000011.10  
**Description** Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly  
**Molecule type** dna  
**Query Length** 21953

**Subject ID** NG\_007264.1  
**Description** Homo sapiens X-linked inhibitor of apoptosis (XIAP), RefSeqGene (LRG\_19) on chromosome X  
**Molecule type** nucleic acid  
**Subject Length** 60775  
**Program** BLASTN 2.6.0+ [Citation](#)

Other reports: [Search Summary](#) [MSA viewer](#)

**Graphic Summary**

Distribution of the top 101 Blast Hits on 1 subject sequences

Mouse over to see the title, click to show alignments

**Color key for alignment scores**

- <40
- 40-50
- 50-80
- 80-200
- >=200

Query  
 102317450 102321449 102325449 102329449 102333449  
 102337449

**Dot Matrix View**

Plot of |c|Query\_42745 vs NG\_007264.1

**Descriptions**

Sequences producing significant alignments:

Select: All None Selected:0

Alignments [Download](#) [GenBank](#) [Graphics](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Homo sapiens X-linked inhibitor of apoptosis (XIAP), RefSeqGene (LRG_19) on chromosome X	403	2.276e+05	20%	5e-113	90%	NG_007264.1



HERV y VIH-1, dando como resultado también positivo secuencias homólogas al VIH-1 y a HERV, en concreto a las LTRs<sup>3</sup> del VIH-1. Ya que se ha encontrado este tipo de homología bastante interesantes, se tratará más adelante en un apartado.

#### 4.1.2 GRIA3 (Glutamate Ionotropic Rceptor AMPA type subunit 3)

El gen GRIA3, codifica una proteína llamada receptor de glutamato 3. Los receptores de glutamato son neurotransmisores excitadores predominantes en el cerebro de los mamíferos y se activan en una variedad de procesos neurofisiológicos normales. La subunidad codificada por este gen pertenece a una familia de receptores de glutamato y está sometida a la edición del pre-ARNm<sup>(23)</sup>.

Como se acaba de comentar, el glutamato o L-glutamato es un aminoácido no esencial y nos puede resultar familiar ya que es empleado en la alimentación como potenciador del sabor en muchos alimentos (glutamato monosódico)<sup>(24)</sup>.

La localización de este gen se encuentra en el cromosoma X, en la posición: Xq25, al igual que XIAP, pero se localiza entre los pares de bases: 123.184.243 a 123.490.915, según la anotación GRch38.p7. El número de acceso en GenBank es: NG\_009377.2.

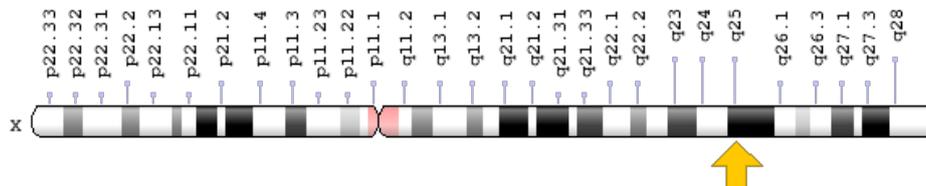


Ilustración 15: Localización del gen GRIA3 en el cromosoma X

A continuación, se muestra una parte del alineamiento del gen BIRC3 (*query*) con respecto al gen GRIA3 (*subject*). BLASTN ha devuelto más de 400 rangos, por lo que el alineamiento completo se puede ver aquí: <https://goo.gl/09JWSe> . Se muestran los resultados y primer alineamiento con mejor *E-value*:

<sup>3</sup> LTR (Long Terminal Repeat o Repetición Terminal Larga): es una secuencia de nucleótidos característica que se encuentra en cada extremo de un elemento retroviral que ha sido integrado en el genoma hospedador. Está implicada en el proceso de integración.



Observamos que la longitud del gen BIRC3 es 21.954 bp y la longitud del gen GRIA3 es de 313.673 bp, obteniendo unos valores de  $E\text{-value} = 0.0$ , Identity = 92% y un Cover de 19%. El alineamiento muestra repeticiones de nucleótidos a lo largo del gen GRIA3 con respecto a BIRC3, lo que indica que sí hay homología entre ambos genes.

Se puede acotar el primer alineamiento haciendo clic en “GenBank” en el primer rango y nos muestra la secuencia, según podemos ver en la siguiente imagen:

### Homo sapiens glutamate ionotropic receptor AMPA type subunit 3 (GRIA3), RefSeqGene on chromosome X

NCBI Reference Sequence: NG\_009377.2

[GenBank](#) [Graphics](#)

```
>NG_009377.2:200377-200820 Homo sapiens glutamate ionotropic receptor AMPA type
subunit 3 (GRIA3), RefSeqGene on chromosome X
TGTTTTTTTTTTTTTTTTGTAATAATTTATTTTTTATTATACTTTAAGTTTATAGGTACATGTGCAC
AATGTGCAGGTTGTTACATATGTATACATGTGCCATGTGGTGTGCTGCACCCATTAACCTCGTCATTTA
GCATTAGGTATATCTCTACTGCTATCCCTCCCCCACCACCAACAGGCCCGGTGTGTGATG
TTCCCTTCCTGTGCCATGTGTTCTCATTGTTCAATTCACCTATGAGTGAGAATATGCGGTGTTTGG
TTTTCTGCTTTGTGATAGTTTGCTCAGAATGATGGTTCCAGCTTCATCCATGTCCTACAAAGGACAT
GAACTCATCCTTTTTTATGGCTGCATAGTATCCATGGTGATATGTTGCACATTTTCTTAATCCAGTCT
ATCACTGATGGACATTTGGGTTGG
```

Como se observa, la cadena es complementaria al alineamiento.

#### 4.1.3 DOCK7 (Dedicator of Cytokinesis 7).

La proteína codificada por el gen DOCK7, consiste en un factor de intercambio de nucleótidos de guanina (GEF) que tiene como papel en la formación de axones y polarización neuronal. Está asociado a epilepsia temprana infantil, problemas de aprendizaje y ceguera.

Se localiza en el cromosoma 1, concretamente en 1p31.3 y la secuencia está comprendida entre los pares de bases: 62.454.716 a 21.688.368, según la última anotación GRCh38.p7<sup>(25)</sup>. El número de acceso del gen en GenBank es: NG\_033073.1.

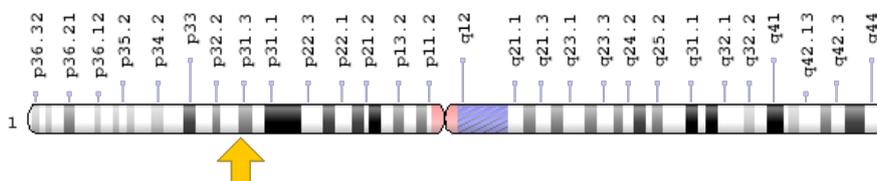


Ilustración 17: Localización de gen DOCK7 en el cromosoma 1

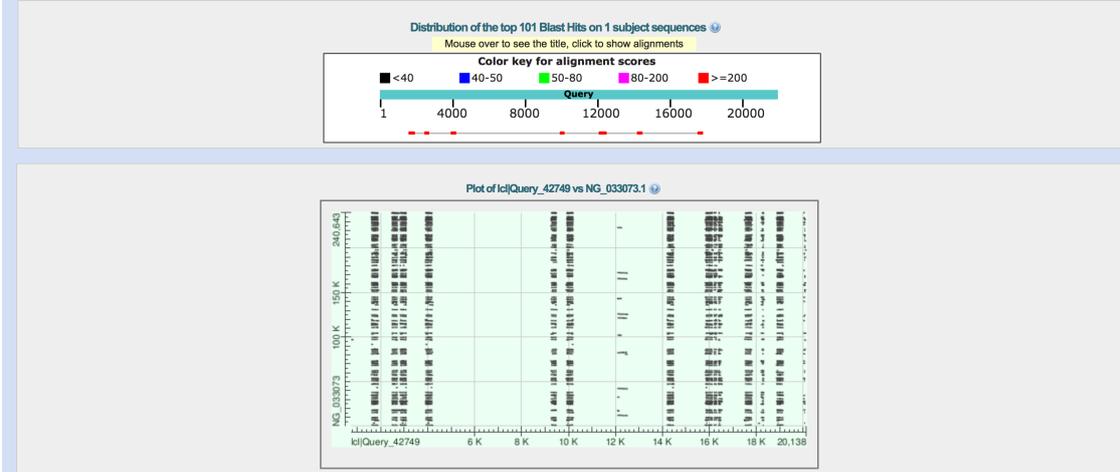
A continuación, se muestra el alineamiento del gen BIRC3 (*query*) con respecto al gen DOCK7 (*subject*). BLASTN ha dado como resultado más de 400 rangos, se puede consultar el alineamiento completo en el siguiente enlace: <https://goo.gl/VmB2HG>. Se muestran a continuación los resultados y el primer alineamiento con mejor  $E\text{-value}$ :

**Query ID** |cljQuery\_42749  
**Description** |NC\_000011.10:102317450-102339403 Homo sapiens chromosome 11, GRCh38.p7 Primary Assembly  
**Molecule type** |nucleic acid  
**Query Length** |21954

**Subject ID** |NG\_033073.1  
**Description** |Homo sapiens dedicator of cytokinesis 7 (DOCK7), RefSeqGene on chromosome 1  
**Molecule type** |nucleic acid  
**Subject Length** |240643  
**Program** |BLASTN 2.6.0+ > Citation

Other reports: > Search Summary [MSA viewer]

**Graphic Summary**



**Descriptions**

Sequences producing significant alignments:

Select: All None Selected:1

Alignments Download GenBank Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
✓ Homo sapiens dedicator of cytokinesis 7 (DOCK7), RefSeqGene on chromosome 1	628	2.802e+05	20%	3e-180	92%	NG_033073.1

Download GenBank Graphics Sort by: E value

Homo sapiens dedicator of cytokinesis 7 (DOCK7), RefSeqGene on chromosome 1  
 Sequence ID: NG\_033073.1 Length: 240643 Number of Matches: 2179

Range 1: 11802 to 12243 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
628 bits(696)	3e-180	408/443(92%)	5/443(1%)	Plus/Minus
Query 12080	GAGCCAACCCAAATGTCATCAGTGATAGACTGGATTAAGAAAATGTGGCACATACACAC	12139		
Sbjct 12243	GAACCAACCCAAATGTCATCAATGATAGACTGGATGAAGAAAATGTGGCACATATACAC	12184		
Query 12140	CATGGAATACATGCGAGCCATAAAAA-GGATGAGTTTCATGTCCTTTATAGGACATGGAT	12198		
Sbjct 12183	CATGGAATACATGCGAGCCATAAAAAAGGATGAGTTTCATGTCCTTTGATAGGACATGGAT	12124		
Query 12199	GAAGCTGGAAACCATCATTTCTGAGCAACTATCACAAGAACAGAAAACCAACACCGCAT	12258		
Sbjct 12123	GAAGCTGGAAACCATCATTTCTGAGCAACTATCACAAGAACAGAAAACCAACACCCACAT	12064		
Query 12259	GTTCTCACTCATAGTTGGGAATTGAACAATGAGAACACTTGGACATAGGAAGGGGAACAT	12318		
Sbjct 12063	GTTCTCACTCATAGTTGGGAATTGAACAATGAGAACATTTGGACACTGGAAGGGGAACAT	12004		
Query 12319	CACACA-CAGGGCTGTTGTTGGGGTAGGGGATGTTGGAGGGATAGCATTAGGAGATACA	12377		
Sbjct 12003	CACACACCAGGGCTGTTGTTGGGGTAGGGGAGGGGAGGGATAGCATTAGGAGATATA	11944		
Query 12378	CCTAATGTTAAATGACGAGTTAACGGGTGCAGCACCAACATGGCAGATGTATACATAT	12437		
Sbjct 11943	CCTACTG-TAAATGACGAGTTAATGGGTGCAGCACCAACATGGCAGATGTATACATAT	11885		
Query 12438	GTAACAACCTGCACATTTATGCACATGTACCCTAGAACTTAAAGTAAAAA--AA'TTaaa	12495		
Sbjct 11884	GTAACAACCTGCACCTTTATGCACATGTACCCTAGAACTTAAAGTATAAATAAAAAAATAA	11825		
Query 12496	aaaaggaaaaaaaaGAATACAGAA	12518		
Sbjct 11824	AAAAATAAAATAAAAAATACAAAA	11802		

Ilustración 18: Resultado de BLASTN del gen BIRC3 y DOCK7

Se observa que la longitud del gen BIRC3 es 21.954 bp y la longitud del gen DOCK7 es de 240.643 bp, obteniendo unos valores de *E-value* de 3e-180, *Identity* de 92% y un *Cover* de 20%. El alineamiento muestra repeticiones de



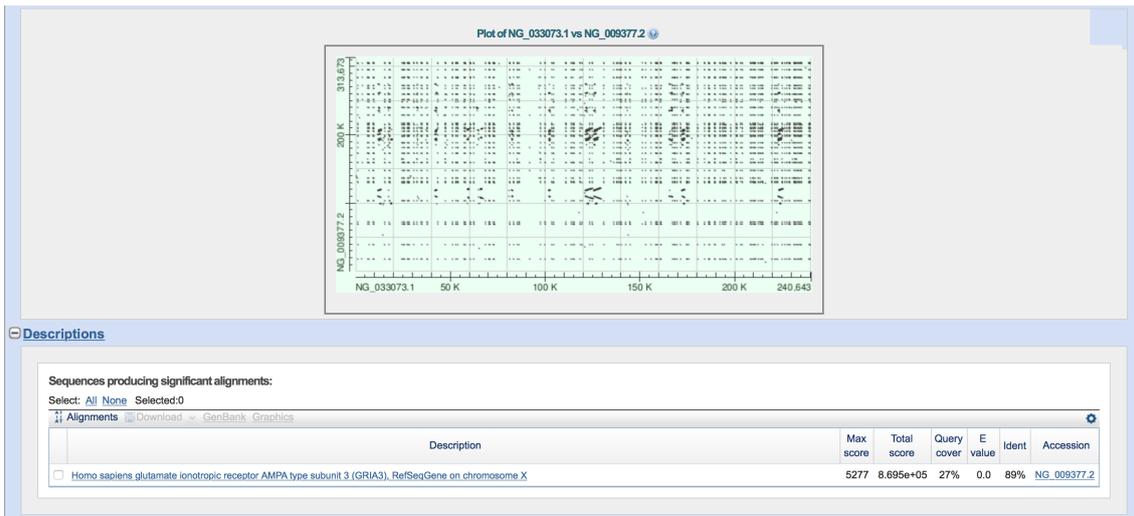


Ilustración 19: Resultado de BLASTN entre los genes GRIA3 y DOCK7

El alineamiento completo se muestra en el siguiente enlace: <https://goo.gl/MLVzvm>.

Como observación, destacamos que el gen DOCK7 se encuentra en el cromosoma 1 y el gen GRIA3 en el cromosoma X, compartiendo largas cadenas de nucleótidos, pero con funciones totalmente distintas, lo que indica que dentro de un gen hay regiones de nucleótidos que son homólogos, pero no necesariamente tienen que cumplir una misma función, los nucleótidos que se encuentran antes o después o incluso los que son eliminados por splicing hacen que éstos desempeñen una función específica.

Podemos concluir que, aunque existe una homología en los cuatro genes BIRC3, XIAP, GRIA3 y DOCK7, cuando obtenemos el ARNm, podemos obtener cadenas de aminoácidos totalmente distintas, ya que como se observa en los alineamientos, existen ajustes en el alineamiento y “gaps”, que son “huecos” en los cuales no hay coincidencia de nucleótido, esto influye a la hora de obtener la proteína final.

En la siguiente imagen, podemos ver un resumen del alineamiento de los 4 genes, por un lado, se observa el gen BIRC3 (*query*) y por otros los genes: XIAP, GRIA3 y DOCK7 (*subject*). El alineamiento completo de estos 4 genes podemos verlo en el siguiente enlace: <https://goo.gl/MLVzvm>

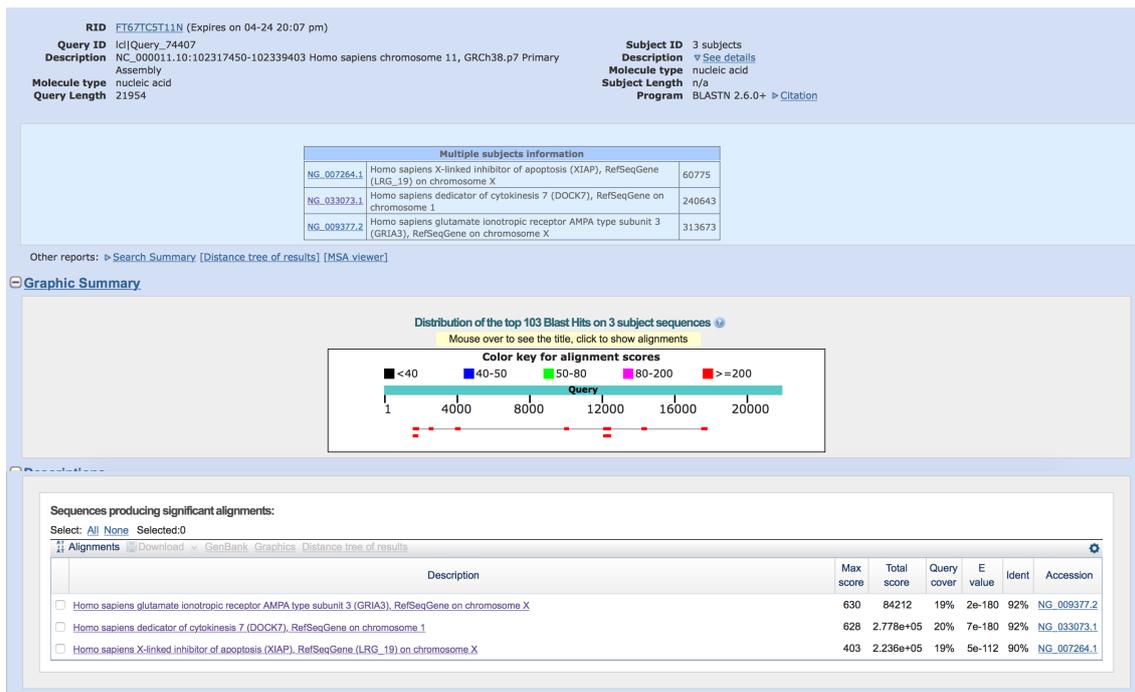


Ilustración 20: Resultado de BLASTN entre los genes: BIRC3, XIAP, GRIA3 y DOCK7

## 4.2. Homologías y similitudes del gen POT1

Para el estudio de la homología, se utiliza como referencia la secuencia de ADN según la anotación original GenBank, acotando los pares de bases que codifican este gen. La anotación es la siguiente: NC\_000007.14:c124929983-124822386 Homo sapiens chromosome 7.

Durante la búsqueda de información de secuencias homólogas al gen POT1, se ha realizado la consulta en las diversas bases de datos genéticas utilizando los algoritmos BLASTN, BLASTX y TBLASTN. A continuación, se comentan los resultados encontrados.

### 4.2.1 SCP2 (Sterol Carrier Protein 2)

El gen SCP2, codifica dos proteínas, la proteína transportadora de estero<sup>4</sup> X (SPCx) y la proteína transportadora de estero<sup>2</sup> (SCP2). Este gen está expresado en órganos implicados en el metabolismo de los lípidos.

El gen SCP2, encuentra localizado en el cromosoma 1, concretamente en 1p32.3 y abarca los pares de bases que van desde: 52.927.229 hasta 53.051.617, según la anotación GRCh38.p7.

<sup>4</sup> El estero<sup>l</sup> es un tipo de esteroide, siendo el más común entre los mamíferos el colesterol.

Para el estudio de este gen se utilizará el número de acceso y región siguiente: NG\_012211.1:4954-129342 Homo sapiens sterol carrier protein 2 (SCP2)<sup>(26)</sup>.

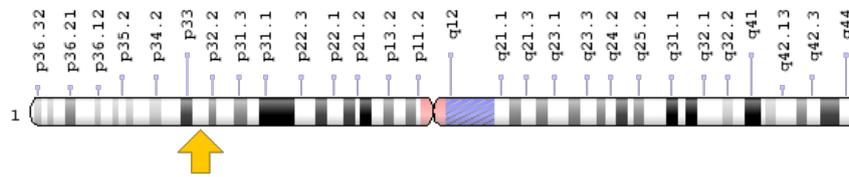


Ilustración 21: Localización del gen SCP2 en el cromosoma 1

A continuación, se muestra el alineamiento del gen POT1 (*query*) con una longitud de 107.598 bp con respecto al gen SCP2 con una longitud de 124.389 bp.

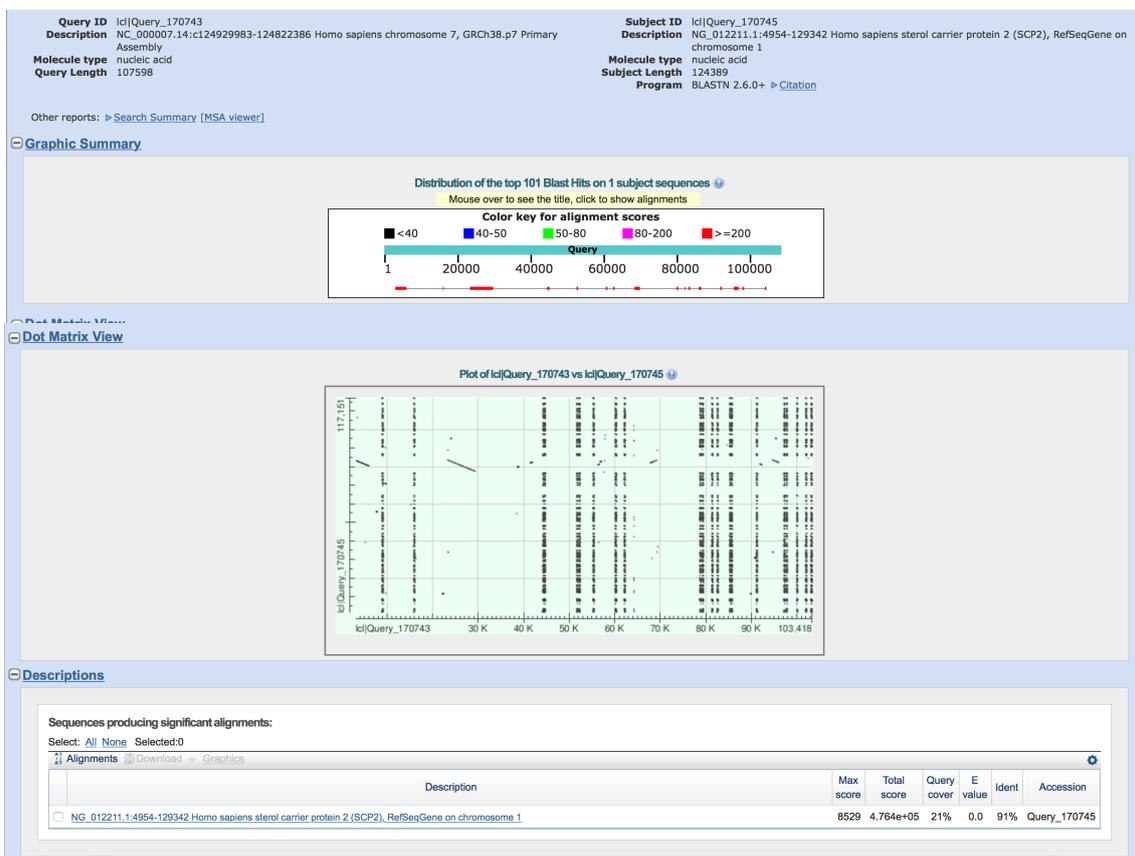


Ilustración 22: Resultado de BLASTN entre el gen POT1 y SCP2

Como resultado, el alineamiento devuelve cientos de rangos como observamos en la matriz de puntos, por lo cual, al ser un alineamiento muy extenso, se puede consultar en la siguiente dirección: <https://goo.gl/4IUbnr>

En los resultados del alineamiento podemos observar lo siguiente: un valor de *E-value* de 0.0, con un Identity igual al 91% y con un Cover de 21%. El alineamiento nos indica que existe homología entre los dos genes en partes a lo largo de los mismos. Es posible afinar un poco más la búsqueda para buscar

alguna proteína que ambos genes tengan en común, para ello, se utilizará la herramienta TBLASTN. Se ha tomado como ejemplo el primer rango de alineación (hacemos clic en GenBank en el primer rango), devolviendo el siguiente resultado:

### Homo sapiens sterol carrier protein 2 (SCP2), RefSeqGene on chromosome 1

NCBI Reference Sequence: NG\_012211.1

[FASTA](#) [Graphics](#)

Go to:

LOCUS NG\_012211 6128 bp DNA linear PRI 31-MAR-2017  
 DEFINITION Homo sapiens sterol carrier protein 2 (SCP2), RefSeqGene on chromosome 1.  
 ACCESSION [NG\\_012211](#) REGION: 82385..88512  
 VERSION NG\_012211.1  
 KEYWORDS RefSeq; RefSeqGene.  
 SOURCE Homo sapiens (human)  
 ORGANISM [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 6128)

Una vez se ha obtenido la secuencia en formato FASTA, se introducirá la consulta en *query* y se realiza la búsqueda en la base de datos: Nucleotide Collection (nr/nt), devolviendo los siguientes resultados:

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> hypothetical protein (L1H 3' region) - human	994	2392	62%	0.0	93%	B34087
<input type="checkbox"/> unknown [Homo sapiens]	974	2350	62%	0.0	91%	<a href="#">AAD38785.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	972	2337	62%	0.0	91%	<a href="#">AAC51279.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	970	2341	62%	0.0	90%	<a href="#">AAC51267.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	969	2339	62%	0.0	90%	<a href="#">AAC51264.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	969	2341	62%	0.0	91%	<a href="#">AAC51261.1</a>
<input type="checkbox"/> RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease	969	2345	62%	0.0	90%	<a href="#">O00370.1</a>
<input type="checkbox"/> unknown [Homo sapiens]	969	2346	62%	0.0	90%	<a href="#">AAD39215.1</a>
<input type="checkbox"/> unknown protein [Homo sapiens]	968	2174	58%	0.0	90%	<a href="#">AA88037.1</a>
<input type="checkbox"/> ORF2 contains a reverse transcriptase domain [Homo sapiens]	968	2340	62%	0.0	90%	<a href="#">AAB59368.1</a>
<input type="checkbox"/> reverse transcriptase homolog - human transposon L1.1	968	2341	62%	0.0	90%	<a href="#">S65824</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	967	2336	62%	0.0	90%	<a href="#">AAC51276.1</a>
<input type="checkbox"/> unknown [Homo sapiens]	967	2342	62%	0.0	90%	<a href="#">AAL50637.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	966	2329	62%	0.0	90%	<a href="#">AAC51263.1</a>
<input type="checkbox"/> reverse transcriptase homolog - human retrotransposon L1	965	2334	62%	0.0	90%	<a href="#">B8888</a>

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> hypothetical protein (L1H 3' region) - human	994	2392	62%	0.0	93%	B34087
<input type="checkbox"/> unknown [Homo sapiens]	974	2350	62%	0.0	91%	<a href="#">AAD38785.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	972	2337	62%	0.0	91%	<a href="#">AAC51279.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	970	2341	62%	0.0	90%	<a href="#">AAC51267.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	969	2339	62%	0.0	90%	<a href="#">AAC51264.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	969	2341	62%	0.0	91%	<a href="#">AAC51261.1</a>
<input type="checkbox"/> RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease	969	2345	62%	0.0	90%	<a href="#">O00370.1</a>
<input type="checkbox"/> unknown [Homo sapiens]	969	2346	62%	0.0	90%	<a href="#">AAD39215.1</a>
<input type="checkbox"/> unknown protein [Homo sapiens]	968	2174	58%	0.0	90%	<a href="#">AA88037.1</a>
<input type="checkbox"/> ORF2 contains a reverse transcriptase domain [Homo sapiens]	968	2340	62%	0.0	90%	<a href="#">AAB59368.1</a>
<input type="checkbox"/> reverse transcriptase homolog - human transposon L1.1	968	2341	62%	0.0	90%	<a href="#">S65824</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	967	2336	62%	0.0	90%	<a href="#">AAC51276.1</a>
<input type="checkbox"/> unknown [Homo sapiens]	967	2342	62%	0.0	90%	<a href="#">AAL50637.1</a>
<input type="checkbox"/> putative p150 [Homo sapiens]	966	2329	62%	0.0	90%	<a href="#">AAC51263.1</a>
<input type="checkbox"/> reverse transcriptase homolog - human retrotransposon L1	965	2334	62%	0.0	90%	<a href="#">B8888</a>

Como se observa, los primeros alineamientos devuelven resultados “hipotéticos” o “putativos”, lo que indica que tenemos poca información acerca de su

procedencia. Por lo que se ha optado a elegir: “ORF2 contains a reverse transcriptase domain [Homo sapiens]”, que se puede consultar en GenBank con el número de acceso: AAB59368.1. Observando la descripción, pertenece a un estudio llamado: “A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome.”<sup>(27)</sup>.

Este gen, está implicado en los centrómeros del cromosoma Y, y como curiosidad, al tratarse de una transcriptasa inversa, se ha buscado similitud utilizando TBLASTN con el virus del VIH-1, devolviendo resultados positivos que serán comentados más adelante.

En conclusión, los genes POT1 y SCP2 son homólogos teniendo en común la secuencia anteriormente comentada.

#### 4.2.2 PPEF1 (Protein Phosphatase with EF-hand Domain 1)

El gen PPEF1, codifica un miembro de la proteína fosfatasa serina / treonina. Esta proteína contiene un dominio de la proteína fosfatasa y aunque su funcionamiento es todavía desconocido, se cree que desempeña un papel específico en la función y/o desarrollo específico de las neuronas sensoriales.

PPEF1 está localizado en el cromosoma X en la posición: Xp22.13 y abarca los pares de bases que van desde: 18675909 hasta 18827921, según la anotación GRCh28.p7. Para el estudio utilizaremos el número de acceso en GenBank: NG\_012519.1<sup>(28)</sup>.



Ilustración 23: Localización del gen PPEF1 en el cromosoma X

A continuación, se observa el resultado del alineamiento utilizando el algoritmo BLASTN. Por un lado, tenemos en *query* la secuencia que codifica el gen POT1 con una longitud de 107598 bp y por otro lado está el gen PPEF1 con una longitud de 124.388 bp.

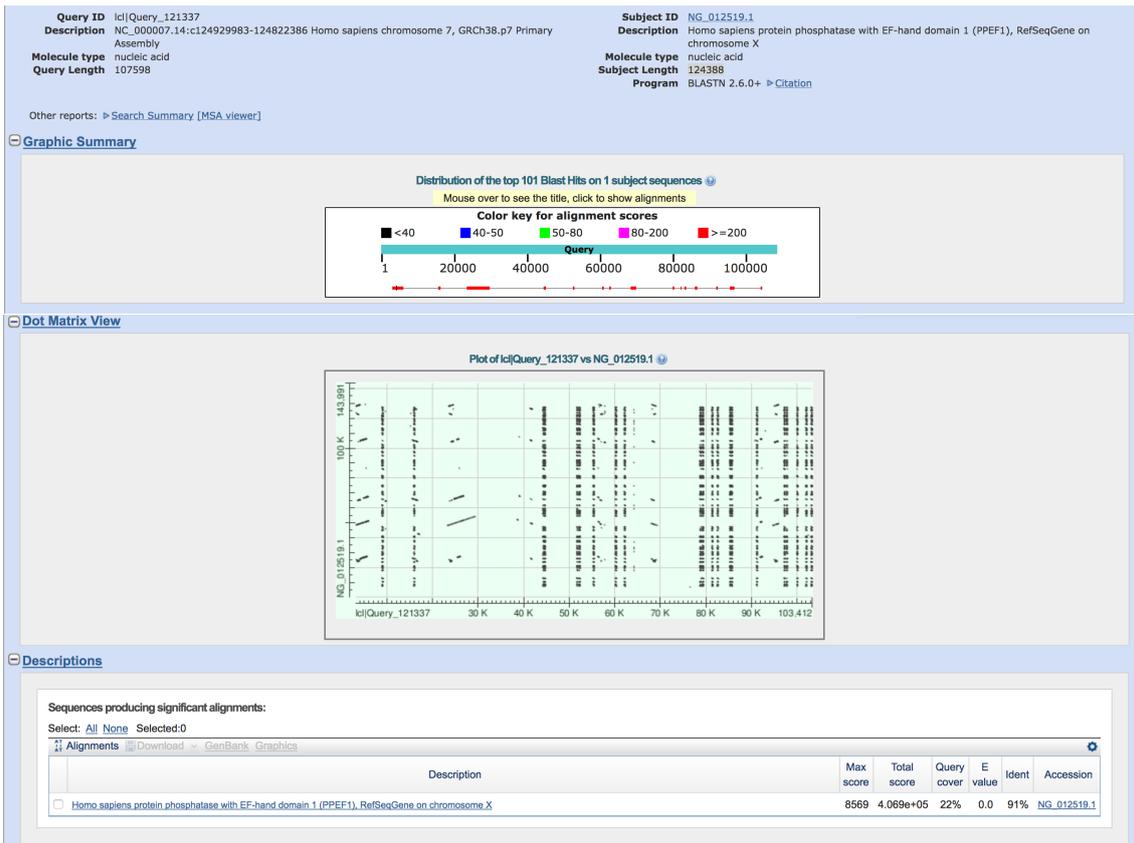


Ilustración 24: Resultado BLASTN entre el gen POT1 y PPEF1

Como se observa en la matriz de puntos, el alineamiento devuelve multitud de rangos, por lo que podemos consultar el alineamiento completo en la siguiente dirección: <https://goo.gl/ouycsc>.

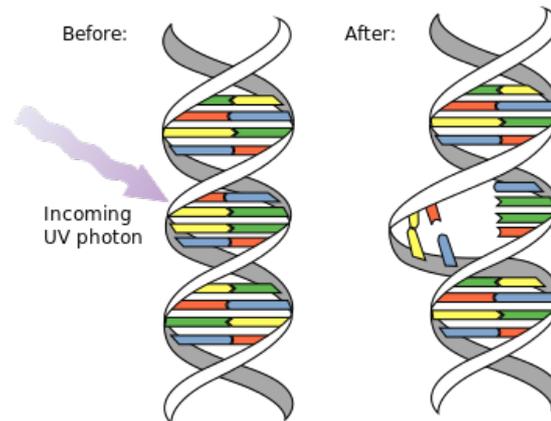
El resultado del alineamiento devuelve un valor *E-value* de 0.0, un *Identity* del 91% y un *Cover* de 22%, lo que indica que existe una homología entre ambos genes a lo largo de la secuencia de ambos genes.

## 5. Mutaciones genéticas

A lo largo de nuestro estudio, se ha hablado de las mutaciones de los genes BIRC3 y POT1 como posibles causas asociadas a la LLC. La búsqueda de genes homólogos ofrece la posibilidad de ampliar el campo de búsqueda para la posible prevención y explicación de casos de LLC, pero el estudio de las mutaciones también es importante para poder comprender las causas de las mismas.

Muchas mutaciones han servido para la evolución en el ser humano o en otras especies, el problema se da cuando esas mutaciones conllevan a una codificación defectuosa de una proteína que lleva a desarrollar un cáncer.

Por ese motivo, se entiende como mutación cuando dada una secuencia de nucleótidos (ADN) existe un cambio o variación con respecto a la secuencia original, es decir, se ha producido una alteración de un gen. Esto es debido a varios factores como la recombinación entre genes, que el gen se transmitió por herencia ya mutado o bien puede ocurrir de una forma espontánea o por agentes mutágenos<sup>5</sup>, en tal caso, el gen mutado impide codificar la proteína correctamente.



*Ilustración 25: Ejemplo de mutación de una cadena de ADN por rayos UV.*

La búsqueda de mutaciones se realiza secuenciando genes de personas con una determinada enfermedad y comparando después con una secuencia consorcio, en el caso del genoma humano, nos basaremos en las secuencias del “Genome Reference Consortium” o “Consortio de Referencia del Genoma”<sup>(29)</sup>. La secuencia consorcio más actualizada viene anotada como GRch38.

Para la búsqueda de mutaciones, se ha utilizado las herramientas y bases de datos disponibles en “International Cancer Genome Consortium”<sup>6</sup>, en concreto “ICGC Data Portal”<sup>7</sup>. Este portal ofrece una gran variedad de base de datos de genes asociados a cánceres y provee de una herramienta de navegación en donde se observa qué nucleótido ha mutado con respecto a la secuencia consorcio.

<sup>4</sup> Cualquier cosa que causa una mutación (cambio en el ADN de una célula). Los cambios que los mutágenos causan en el ADN pueden dañar las células y provocar una enfermedad, como el cáncer. Entre los ejemplos de mutágenos están las sustancias radiactivas, los rayos x, la radiación ultravioleta y ciertas sustancias químicas. Fuente: <https://www.cancer.gov/espanol/publicaciones/diccionario?crid=601170>

<sup>6</sup> <https://icgc.org>

<sup>7</sup> <https://dcc.icgc.org>

## 5.1 Mutaciones del gen BIRC3

Accediendo al portal ICGC Data, se realiza una búsqueda del gen BIRC3 obteniendo como resultado lo siguiente:



**ICGC Data Portal**

KEYWORD SEARCH

Q BIRC3

66 results found for BIRC3

Gene Donor Mutation Project GeneSet File Compound

**BIRC3**  
Gene

**Name:** baculoviral IAP repeat containing 3  
**Synonyms:** API2, c-IAP2, cIAP2, hiap-1, MALT2, MIHC, RNF49  
**UniProtKB/Swiss-Prot:** Q13489  
**Ensembl ID:** ENSG0000023445

Como se aprecia en la imagen anterior, la búsqueda nos devuelve 66 resultados para el gen BIRC3, centrándose en el Gen (Gene en la imagen) para proceder al estudio completo. También se encuentra el acceso a la secuencia en la base de datos UniProtKB/Swiss-Prot con el número: Q13489, podemos acceder a la secuencia a través de portal <https://www.ncbi.nlm.nih.gov>

Entrando en el detalle que devuelve la búsqueda del gen BIRC3, la aplicación devuelve información sobre este gen y una gráfica con una tabla de datos de distribución de los cánceres asociados al mismo, como se puede observar en la siguiente imagen:

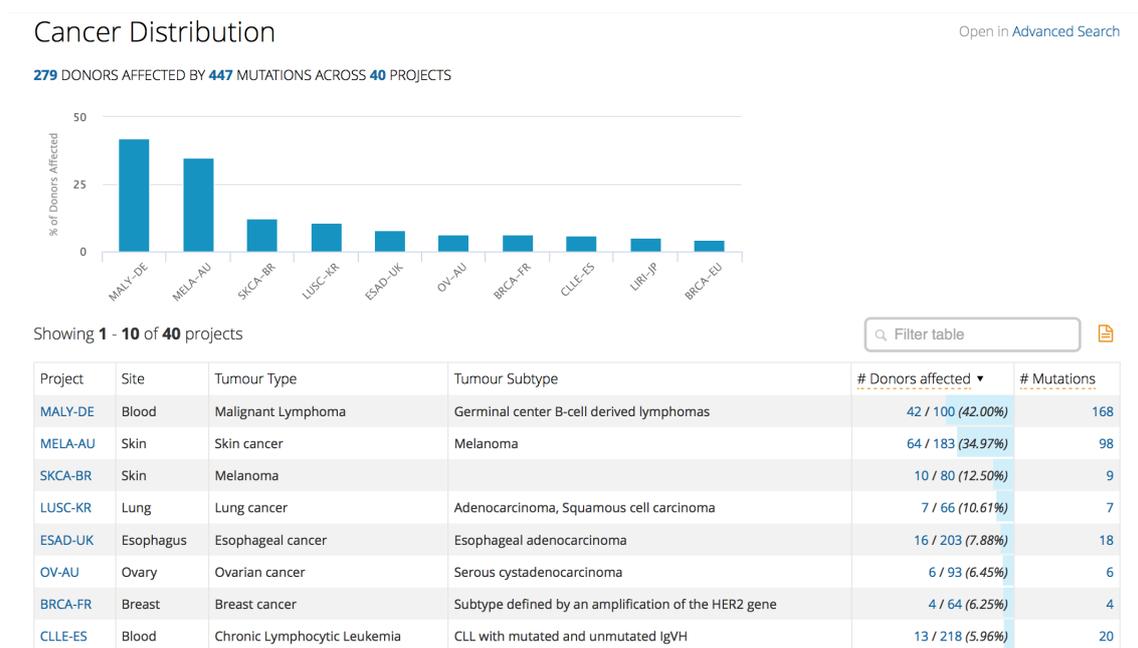


Ilustración 26: Tabla de distribución de cánceres asociados a BIRC3

Aparece reflejado que hay 279 donantes afectados por 447 mutaciones recogidos en 40 proyectos. En concreto, si se observa la tabla, se puede encontrar un proyecto llamado CLLE-ES que corresponde a Chronic Lymphocytic Leukemia (CLL) que corresponde al estudio que se está llevando a cabo. En concreto, la tabla informa que hay 20 mutaciones asociadas a la LLC.

A continuación, se observa la tabla en donde se refleja la siguiente información:

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types:											# Mutations	# Genes
							SSM	CNSM	SISM	SGV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN		
DO6456	CLLE-ES	Blood	Female	76	A	3,246	✓	✓	✓	--	✓	--	✓	✓	--	--	--	2	1
DO51989	CLLE-ES	Blood	Female	61	A	1,931	✓	✓	--	--	--	--	✓	✓	--	--	--	4	1
DO6402	CLLE-ES	Blood	Male	58	A	3,144	✓	--	✓	--	✓	--	✓	✓	--	--	--	1	1
DO52701	CLLE-ES	Blood	Male	57	A	5,908	✓	--	✓	--	--	--	✓	✓	--	--	--	1	1
DO223410	CLLE-ES	Blood	Female	70	A	2,912	✓	✓	✓	--	--	--	✓	✓	--	--	--	3	1
DO6418	CLLE-ES	Blood	Male	62	A	2,057	✓	✓	✓	--	✓	--	✓	✓	--	--	--	1	1
DO6742	CLLE-ES	Blood	Male	62	A	5,042	✓	✓	--	--	✓	--	✓	✓	--	--	--	2	1
DO6376	CLLE-ES	Blood	Male	70	A	4,004	✓	--	--	--	✓	--	✓	✓	--	--	--	1	1
DO6700	CLLE-ES	Blood	Female	63	A	4,119	✓	✓	✓	--	✓	--	✓	✓	--	--	--	1	1
DO223397	CLLE-ES	Blood	Female	70	A	2,350	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO7196	CLLE-ES	Blood	Male	74	A	4,665	✓	✓	--	--	--	--	✓	✓	--	--	--	1	1
DO51966	CLLE-ES	Blood	Female	62	A	5,281	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO6486	CLLE-ES	Blood	Female	71	A	4,595	✓	✓	✓	--	✓	--	✓	✓	--	--	--	1	1

Ilustración 27: Tabla de los pacientes que presentan mutación del gen BIRC3 asociado a la LLC

Se procede a comentar los siguientes datos de interés:

- **ID:** corresponde a la identificación del donante.
- **Project:** El nombre del proyecto de estudio, en nuestro caso es la LLC.
- **Site:** Corresponde a la procedencia de la muestra, en este caso es de la sangre.
- **Gender:** Es el sexo del donante Hombre (male) o Mujer (Female).
- **Age:** Corresponde a la edad del donante.
- **Stage:** Corresponde al estadio del cáncer.
- **Survival days:** Días de supervivencia del donante una vez detectado el cáncer
- **SSM:** Mutación somática simple, es decir, que afecta a las células somáticas del donante. Las células somáticas son las que conforman el crecimiento de los tejidos y órganos de un ser vivo y proceden de las células madre, pero no son células germinales como un gameto (óvulo o espermatozoide).
- **Mutations:** Es el número de mutaciones del donante.
- **Genes:** Es el número de genes extraídos del donante.

Como se observa en la tabla, la edad de los donantes está comprendida entre los 60 y 70 años, edad típica de los enfermos de LLC.

A continuación, se observa la tabla que informa del tipo de mutaciones:

ID	DNA change	Type	Consequences	# Donors affected	
				Total	Across all Projects
MU47468668	chr11:g.102188685A>G	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU53160118	chr11:g.102188632T>C	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU56471599	chr11:g.102189205A>G	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU46849410	chr11:g.102188528T>C	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU50565361	chr11:g.102207657C>-	deletion of <=200bp	Frameshift: <i>BIRC3 Q547</i> Exon: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU52146305	chr11:g.102191451AT>-	deletion of <=200bp	Upstream: <i>BIRC3</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU51509242	chr11:g.102201944>T	insertion of <=200bp	Frameshift: <i>BIRC3 R432S?</i> Downstream: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU45327945	chr11:g.102188343T>C	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU53471001	chr11:g.102188552C>T	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU47470193	chr11:g.102188662A>G	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU56294418	chr11:g.102188616A>C	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU44300198	chr11:g.102205934C>A	single base substitution	Upstream: <i>BIRC3</i> Downstream: <i>BIRC3</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU57012128	chr11:g.102191173A>G	single base substitution	Upstream: <i>BIRC3</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU47466715	chr11:g.102188730A>G	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU50885547	chr11:g.102188690C>T	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU53156070	chr11:g.102188694T>C	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU47469637	chr11:g.102188677A>G	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU53160639	chr11:g.102188421T>A	single base substitution	Downstream: <i>RNU6-952P</i> Intron: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU50560456	chr11:g.102207676->A	insertion of <=200bp	Frameshift: <i>BIRC3 E553E?</i> Exon: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU44303234	chr11:g.102201929AAGGG>-	deletion of <=200bp	Frameshift: <i>BIRC3 IRE427</i> Downstream: <i>BIRC3</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)

Ilustración 28: La tabla informativa sobre el número de mutación y tipo de mutación

La información que ofrece la tabla es la siguiente:

- a) ID: que corresponde al número de mutación.
- b) *DNA change*: esta columna nos ofrece la información más interesante sobre la mutación, ya que nos dice el cromosoma afectado, donde empieza la secuencia en bp, la base correspondiente según el consenso y la mutación causante del gen. Por ejemplo, si observamos el ID MU4768668, vemos que la secuencia corresponde al cromosoma 11, la referencia se localiza en 102.188.685 y la base correspondiente debería ser A y ha mutado a G (A>G).
- c) *Type*: Es el tipo de mutación que se observa:
  - *Single base substitution*: es cuando la mutación consiste sólo en una sustitución de una base por otra.
  - *Deletion of*: corresponde a la delección de una o más bases, es decir, la eliminación de nucleótidos que deberían de estar en la secuencia del gen.
  - *Insertion of*: corresponde a una inserción de nucleótidos donde no deberían insertarse, es decir, la secuencia tiene nucleótidos demás con respecto al consenso.
- d) *Consequences*: Consecuencias causadas por la mutación, estas pueden ser dadas en intrones o en exones.
- e) *Donors affected*: Corresponde el número de donantes afectados y donantes afectados a través de todos los proyectos.

Por último, la herramienta “Genome Viewer” que se encuentra en la sección de mutaciones (ver imagen), ofrece un mapa completo que indica dónde se encuentran las mutaciones comparándolas con la secuencia consenso.

Mutations

 Genome Viewer

Esta información es muy útil a la hora de localizar las mutaciones y el estudio de las mismas.

La siguiente imagen, muestra la secuencia de ADN del gen BIRC3 y las mutaciones asociadas marcadas con un cuadrado verde o amarillo que ofrece información acerca del tipo de mutación:

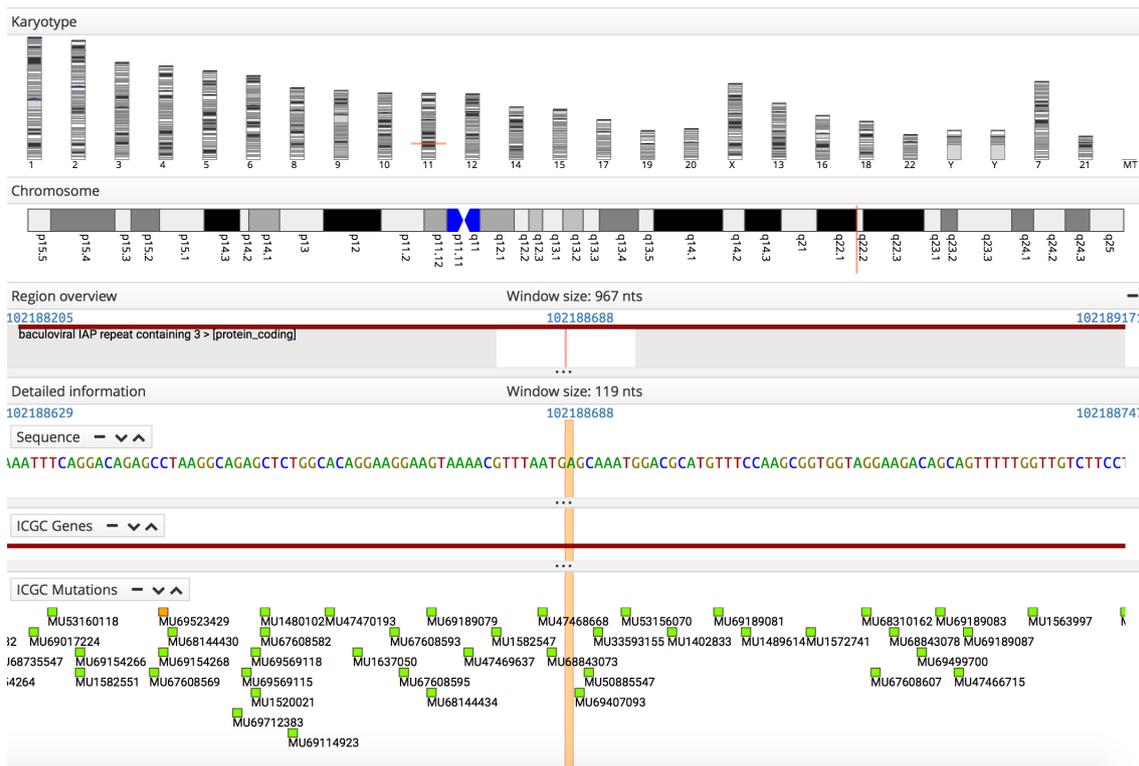


Ilustración 29: Información sobre las distintas mutaciones dentro de una secuencia de ADN

La información que ofrece esta imagen es la siguiente:

- Cariotipo (*Karyotype*): que es el conjunto de cromosomas de la célula. En este caso, aparece marcado el cromosoma 11.
- Cromosoma (*Chromosome*): identificación de la región del cromosoma donde se encuentra el gen:
- Gen de estudio (*Region overview*): el nombre del gen y la proteína que codifica.
- Secuencia detallada (*Detailed information*): es la secuencia de nucleótidos detallada en la región de estudio.
- ICGC Mutaciones (*ICGC Mutations*): son las mutaciones identificadas por MU más un número obtenidas de los donantes.

Como se puede observar de los resultados, es una información muy completa y detallada, pero aún se puede obtener más información pasando el cursor del ratón por encima de los cuadros, como se puede observar a continuación:

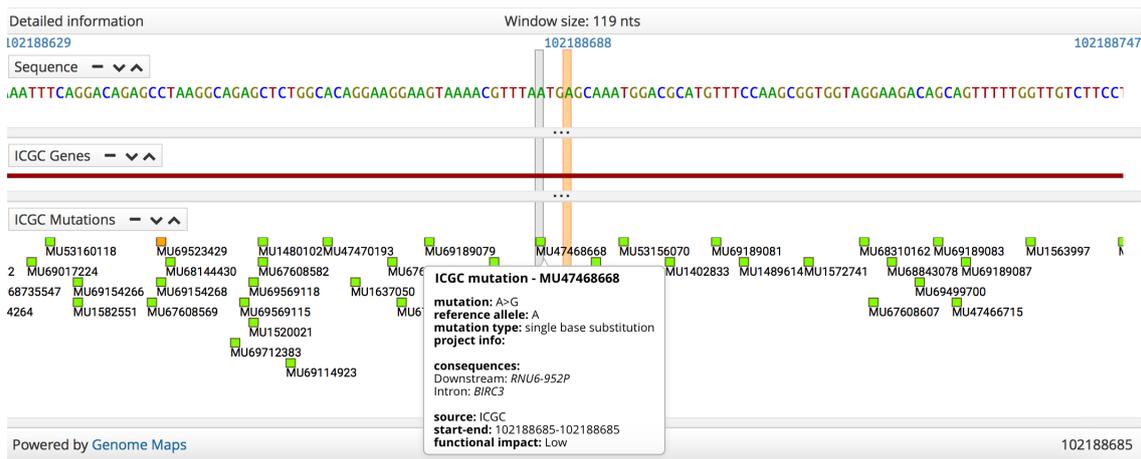


Ilustración 30: Información detallada sobre el tipo de mutación

Pasando el cursor por encima del cuadrado identificado como MU47468668, se obtiene la información anteriormente comentada pero resumida para una comprensión más sencilla.

Como ejemplo, se puede realizar la búsqueda de la mutación MU52146305, seleccionándola de la lista que ofrece la herramienta de navegación y ver con exactitud de qué se trata:

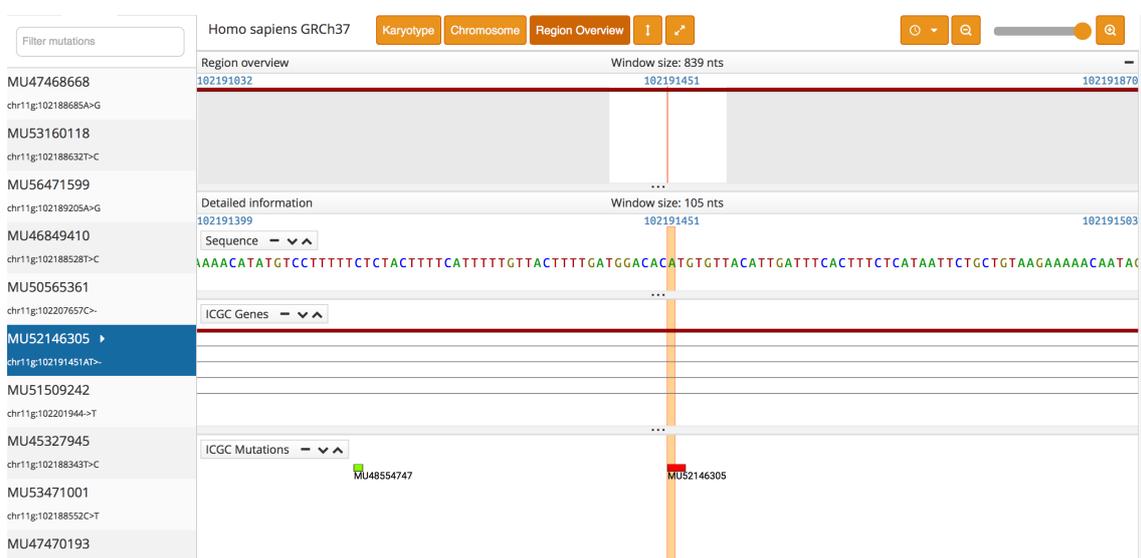


Ilustración 31: Ejemplo de búsqueda de mutaciones en una secuencia

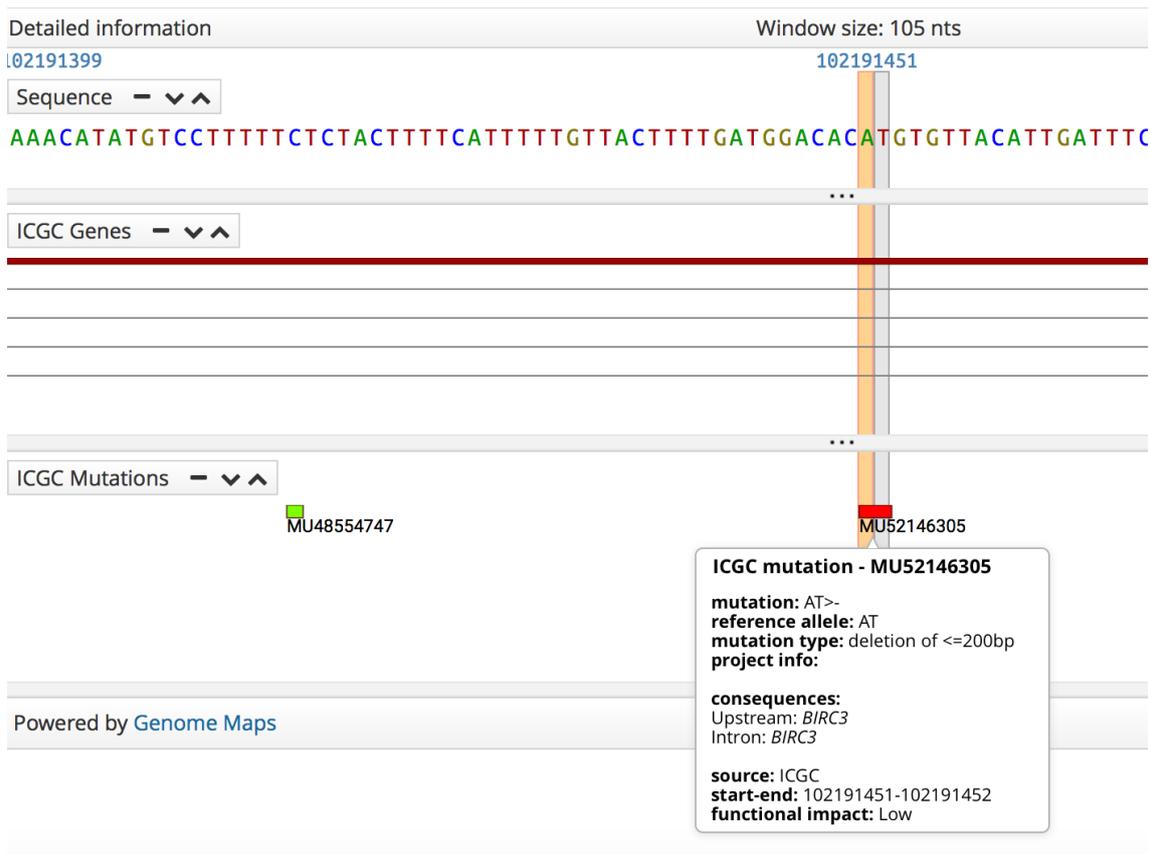


Ilustración 32: Ejemplo de delección de dos nucleótidos en el gen *BIRC3*

Se trata de la delección de dos nucleótidos, en concreto A y T.

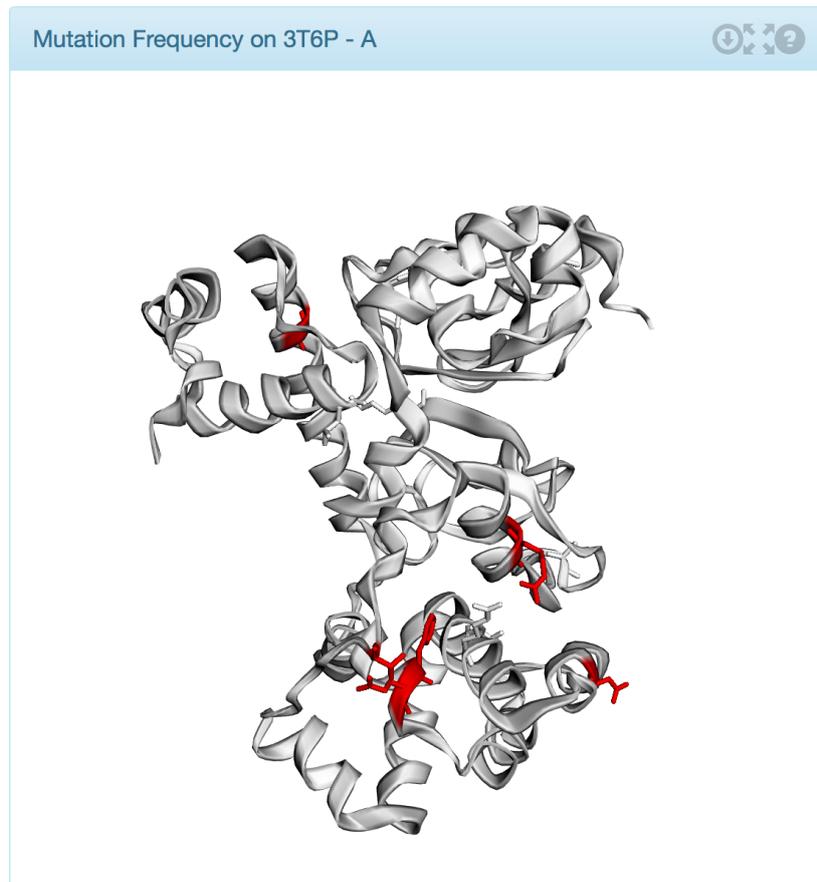
En conclusión, el estudio de la mutación del gen *BIRC3* a través de la herramienta proporcionada por ICGC, ayuda al estudio y al análisis del gen en profundidad, pudiendo ampliar el estudio sobre la población afectada por la LLC e incluso para el desarrollo de fármacos para el tratamiento del cáncer. La identificación de las mutaciones ayuda a la comprensión de la proteína que codifica el gen, es este caso, la proteína codificada es 3T6P-A que es estudiada en el siguiente apartado.

### 5.1.1. Proteína 3T6P-A codificada por el gen *BIRC3*

La proteína 3T6P-A, se encarga de la apoptosis de la célula<sup>(30)</sup>. Si existe una mutación en el gen que se encarga de la codificación de la proteína, puede dar lugar a que ésta no realice la función correcta, por lo que la identificación de las regiones afectadas de la proteína, puede ayudar a la búsqueda de un posible tratamiento para la LLC.

Existe una herramienta que ofrece información sobre la estructura de la proteína y de las regiones afectadas. La herramienta en cuestión se localiza en el sitio web: <http://cancer3d.org>.

Introduciendo el gen BIRC3 en la búsqueda, devuelve como resultado la siguiente imagen:



*Ilustración 33: Proteína 3T6P*

El simulador, devuelve una imagen en 3D en la que se puede observar una escala de color que va de blanco (menor número de mutaciones) a rojo (mayor número de mutaciones) y las posiciones sin mutaciones se muestran en gris, es decir, las regiones en rojo son las más susceptibles a mutaciones.

Para poder estudiar con profundidad qué secuencia de aminoácidos está afectada por mutación, por ejemplo, una mutación de sustitución, se puede descargar el fichero de la proteína y realizar una simulación con la aplicación Chimera<sup>8</sup>. Esta aplicación, ayuda a analizar la proteína y al mismo tiempo se puede comparar con otra proteína consenso para ver las diferencias entre ellas y localizar exactamente dónde se encuentran las mutaciones dentro de la proteína.

Como vemos en la siguiente imagen, hay 3 ventanas:

- La ventana de la izquierda, muestra la proteína de estudio mutada.

<sup>8</sup> <https://www.cgl.ucsf.edu/chimera/>



KEYWORD SEARCH

POT1

99 results found for POT1

- All
- Gene
- Donor
- Mutation
- Project
- GeneSet
- File
- Compound

**POT1**  
Gene

**Name** protection of telomeres 1  
**Synonyms** DKFZp586D211, hPot1  
**UniProtKB/Swiss-Prot** Q9NUX5  
**Ensembl ID** ENSG00000128513

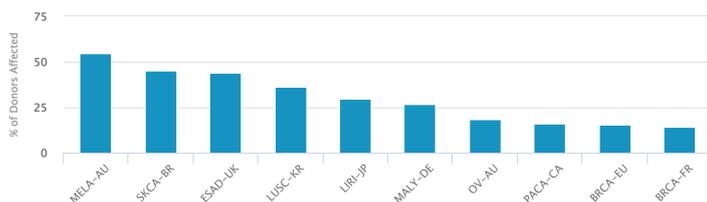
Ilustración 35: Búsqueda en el portal ICGC

La información que reporta es de 99 resultados, aunque se deberá centrar la búsqueda en el primer resultado que devuelve el nombre: “Protection Of Telomeres 1”. Haciendo clic en el nombre del gen, obtenemos 44 proyectos, por lo que se aplica el filtro con las siglas CLL (que corresponden a LLC en inglés) y se obtiene el siguiente resultado:

Cancer Distribution

[Open in Advanced Search](#)

735 DONORS AFFECTED BY 1,216 MUTATIONS ACROSS 44 PROJECTS



Showing 1 projects

CLL

Project	Site	Tumour Type	Tumour Subtype	# Donors affected	# Mutations
CLLE-ES	Blood	Chronic Lymphocytic Leukemia	CLL with mutated and unmutated IgVH	15 / 218 (6.88%)	17

Ilustración 36: Gráfico de distribución de los tipos de cánceres afectados por la mutación del gen POT1

El resultado muestra que se han encontrado 17 mutaciones correspondientes a 15 donantes. La siguiente tabla devuelve información más detallada:

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types:											# Mutations	# Genes
							SSM	CNSM	SISM	SOV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN		
DO223381	CLLE-ES	Blood	Female	76	A	3,612	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO6934	CLLE-ES	Blood	Male	68	B	2,698	✓	✓	✓	--	--	--	✓	--	--	--	--	2	1
DO6898	CLLE-ES	Blood	Female	76	B	1,038	✓	✓	--	--	--	--	--	--	--	--	--	1	1
DO52703	CLLE-ES	Blood	Male	63	A	4,429	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO6410	CLLE-ES	Blood	Male	53	A	2,653	✓	--	✓	--	✓	--	✓	✓	--	--	--	1	1
DO52723	CLLE-ES	Blood	Male	58	A	2,429	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO223398	CLLE-ES	Blood	Male	71	A	2,163	✓	✓	--	--	--	--	✓	✓	--	--	--	1	1
DO223413	CLLE-ES	Blood	Male	63		1,388	✓	✓	--	--	--	--	✓	✓	--	--	--	1	1
DO52709	CLLE-ES	Blood	Female	63	A	6,254	✓	✓	✓	--	--	--	--	--	--	--	--	1	1
DO52720	CLLE-ES	Blood	Female	61	A	3,034	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO6496	CLLE-ES	Blood	Male	44	A	1,696	✓	✓	✓	--	✓	--	✓	✓	--	--	--	1	1
DO51966	CLLE-ES	Blood	Female	62	A	5,281	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO6904	CLLE-ES	Blood	Female	53	A	5,887	✓	✓	--	--	--	--	--	--	--	--	--	1	1
DO6470	CLLE-ES	Blood	Male	69	B	1,630	✓	✓	--	--	✓	--	✓	✓	--	--	--	1	1
DO52725	CLLE-ES	Blood	Female	70	A	3,285	✓	✓	--	--	--	--	✓	✓	--	--	--	2	1

Ilustración 37: Tabla de los pacientes que presentan mutación del gen POT1 asociado a la LLC

Se observa que la media de edad de las personas afectadas por la mutación POT1 está entre los 50 y 70 años, es decir, 10 años por debajo de la media en comparación con el gen BIRC3. Este tramo de edad, entra dentro del rango de desarrollo de la LLC.

La siguiente tabla muestra el tipo de mutaciones asociadas al POT1:

ID	DNA change	Type	Consequences	# Donors affected		Link
				Total	Across all Projects	
MU50282	chr7:g.124537227T>A	single base substitution	Initiator Codon: <b>POT1 M1L</b> Start Gained: <b>POT1</b> Exon: <b>POT1</b> Downstream: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU49640	chr7:g.124537268T>A	single base substitution	Splice Acceptor: <b>POT1</b> Upstream: <b>POT1</b> Downstream: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU54819	chr7:g.124532338A>T	single base substitution	Missense: <b>POT1 Y36N</b> 5 UTR: <b>POT1</b> Exon: <b>POT1</b> Downstream: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU49648	chr7:g.124511022A>C	single base substitution	Stop Gained: <b>POT1 Y66*</b> Exon: <b>POT1</b> Intron: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU57301519	chr7:g.124518002A>G	single base substitution	Intron: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU49773	chr7:g.124499045T>C	single base substitution	Missense: <b>POT1 Y223C, Y20C, Y92C, Y53C</b> 3 UTR: <b>POT1</b> Exon: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link
MU57301939	chr7:g.124517905A>G	single base substitution	Intron: <b>POT1</b>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	Link

<input type="checkbox"/>	MU54599276	chr7:g.124466613T>C	<a href="#">single base substitution</a>	Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU51865781	chr7:g.124464066G>A	<a href="#">single base substitution</a>	Stop Gained: <a href="#">POT1 Q488*</a> , <a href="#">Q117*</a> , <a href="#">Q619*</a> 3 UTR: <a href="#">POT1</a> Exon: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU45987128	chr7:g.124477517T>C	<a href="#">single base substitution</a>	Downstream: <a href="#">POT1</a> Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU50719489	chr7:g.124487266G>T	<a href="#">single base substitution</a>	Upstream: <a href="#">POT1</a> Downstream: <a href="#">POT1</a> Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU54272770	chr7:g.124493143ATTG->	<a href="#">deletion of &lt;=200bp</a>	Frameshift: <a href="#">POT1 SM47</a> , <a href="#">SM119</a> , <a href="#">SM250</a> Exon: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU46522824	chr7:g.124532780T>C	<a href="#">single base substitution</a>	Exon: <a href="#">POT1</a> Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU51518254	chr7:g.124551319A>G	<a href="#">single base substitution</a>	Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU55087158	chr7:g.124512498T>G	<a href="#">single base substitution</a>	Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU54593615	chr7:g.124466615A>T	<a href="#">single base substitution</a>	Intron: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>
<input type="checkbox"/>	MU52972478	chr7:g.124503534C>A	<a href="#">single base substitution</a>	Missense: <a href="#">POT1 W6L</a> , <a href="#">W139L</a> Upstream: <a href="#">POT1</a> 3 UTR: <a href="#">POT1</a> Exon: <a href="#">POT1</a>	1 / 1 (100.00%)	1 / 10,648 (0.01%)	<a href="#">▶</a>	<a href="#">▶</a>

Ilustración 38: La tabla informativa sobre el número de mutación y tipo de mutación

La mayoría de las mutaciones corresponden a un cambio en una base con respecto a la secuencia consenso, sólo hay una mutación que corresponde a una delección. En cuanto a las consecuencias, la mayoría corresponden a delección en intrones o en 3' UTR (UnTranslated Region – Región no Traducida de los genes), que es la región no traducida de los genes. Cabe recordar que las UTRs son muy importantes en la regulación de las expresiones genéticas y están implicadas en la correcta expresión de los genes, ya que hay proteínas que reconocen secuencias específicas que no son codificantes del 3' UTR.

Las siguientes imágenes muestran la siguiente información: el cariotipo del gen POT1, su localización dentro del cromosoma 7, la secuencia de ADN y las mutaciones asociadas marcadas con un cuadrado verde o rojo, ofreciendo información acerca del tipo de mutación. En concreto, se realiza la búsqueda de la mutación asociada a la delección anteriormente comentada que lleva como identificador MU54373770.

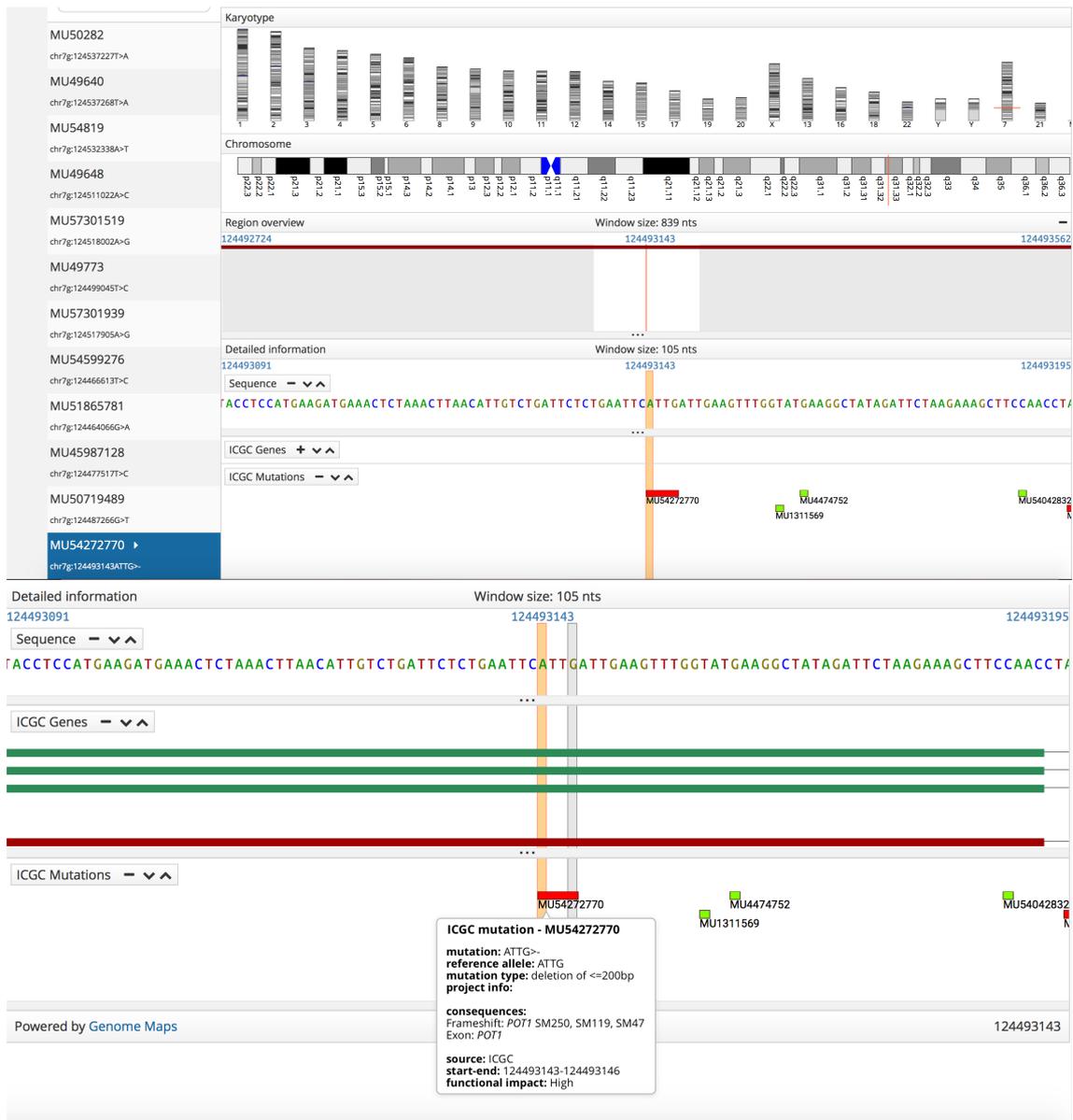


Ilustración 39: Resultado que muestra información sobre el tipo de mutación y la región de ADN afectada

Como se observa en estas imágenes, la mutación consiste en la eliminación de 4 nucleótidos, en concreto ATTG, lo que produce un cambio de marco (frameshift) que es importante a tener en cuenta, ya que tiene un impacto funcional alto.

Un frameshift o cambio de marco, es una mutación con desplazamiento, desfase o cambio de marco de lectura y es un tipo de mutación causada, en este caso, por la delección de un número de nucleótidos que no es múltiplo de tres en la secuencia de ADN.

El gen POT1 codifica la proteína 3KJP <sup>(31)</sup>, las mutaciones anteriormente comentadas pueden afectar a la estructura de la proteína. En el siguiente apartado lo estudiamos con más detenimiento.

### 5.2.1 Proteína 3KJP-A codificada por el gen POT1

La proteína 3KJP-A, se encarga de proteger los telómeros de los cromosomas y se encuentra en los extremos de los brazos de los cromosomas evitando que se adhieran a otros cromosomas, una mutación de este gen provocará que la proteína no actúe correctamente y el proceso de división celular se verá alterado, por lo que el estudio de las mutaciones asociadas a esta proteína puede ayudar a un posible tratamiento de la LLC, como en el caso del gen BIRC3 anteriormente comentado.

La herramienta que nos ofrece “Cancer3d.org”, muestra las regiones de la proteína que están afectadas por las mutaciones:

<http://cancer3d.org/mainView.php?ensemblID=ENSP00000350249&state=MF>

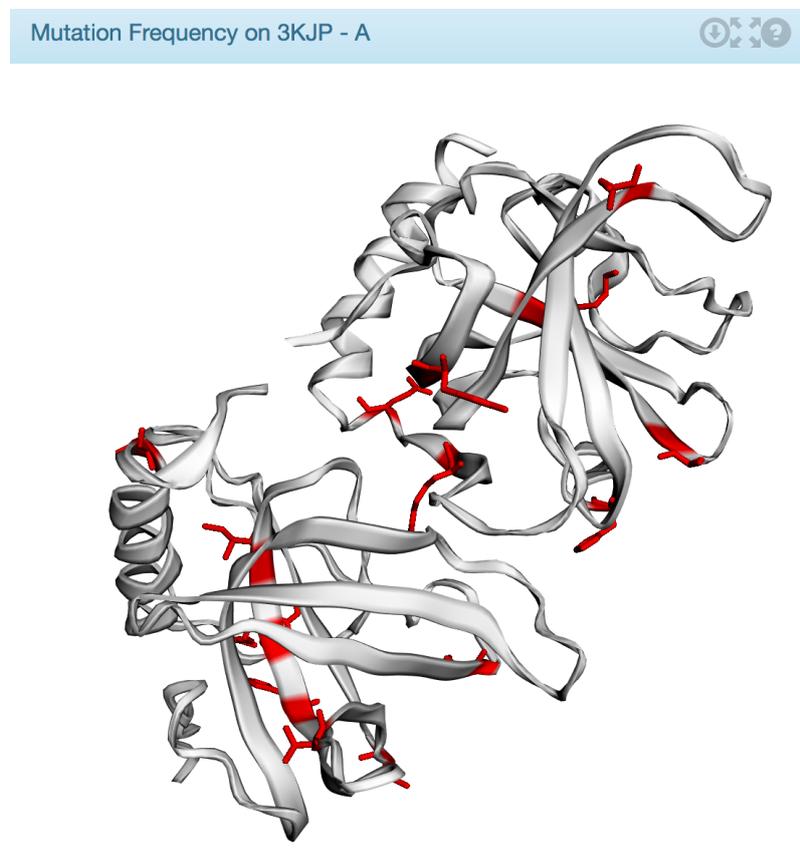


Ilustración 40: Proteína 3KJP-A

Se observan muchas regiones afectadas con respecto a la anterior proteína 3TP6, esto lleva a deducir que la mutación del gen POT1 es mucho más evidente. Se procede a realizar la simulación en Chimera para estudiar las regiones y el tipo de mutación para definir más su estudio.

Para comprender mejor la mutación de la proteína 3JKP, se observa en la siguiente imagen en 3D la proteína sin mutación, según secuencia consenso obtenida de PDB/Uniprot.

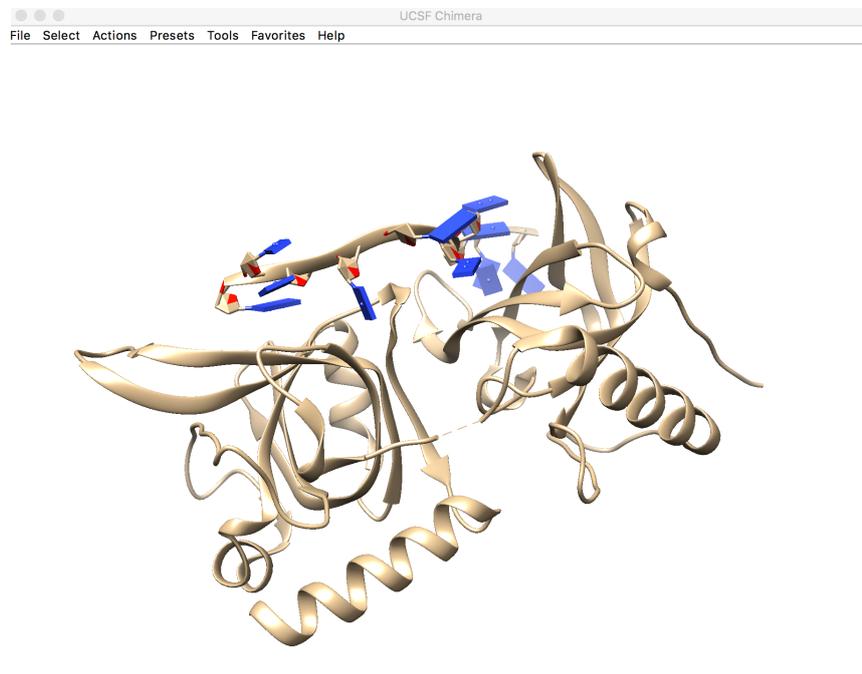


Ilustración 41: Proteína 3JKP consenso.

A continuación, se muestra la Ilustración 42 correspondiente a la proteína 3KJP mutada y la secuencia de aminoácidos y se observa lo siguiente:

- La ventana de la izquierda, muestra la estructura de la proteína en 3D.
- La ventana de la derecha arriba, muestra la secuencia de la proteína mutada 3KJP, marcada en verde.
- La ventana de la derecha abajo, muestra la secuencia completa de la proteína, y marcado en marrón, las secuencias que faltan de aminoácidos según la secuencia consenso. Es decir, la proteína mutada no tiene todos los aminoácidos con respecto a la secuencia consenso.

The screenshot displays the UCSF Chimera interface. On the left, a 3D ribbon model of the protein is shown in light green. On the right, two sequence alignment windows are visible. The top window, titled 'chain A: protection of telomeres protein 1', shows the mutated sequence with green highlights: MS L V P A T N Y I Y T P L N Q L K G G T I V N V Y G V V K F F K P P Y L S K G I D Y C S V V T I V, D Q T N V K L T C L L F S G N Y E A L P I I Y K N G D I V R F H R L K I Q V Y K K E T Q G I T S S G, F A S L T F E G T L G A P I I P R T S S K Y E N F I T E D H K M V E A L R V W A S T H M S P S W I L, K L C D V Q P M Q Y F D L T C Q L L G K A E V D G A S F L L K V W D G T R T P F P S W R V L I Q D, L V L E G D L S H I H R L Q N L T I D I L V Y D N H V H V A R S L K V G S F L R I Y S L H T K L G S, and M N S E N Q T M L S L E F H L H G G T S Y G R G I R V L P E S N S D V D Q L K K D L E S A N L T A. The bottom window shows the full consensus sequence with brown highlights indicating missing amino acids: M S L V P A T N Y I Y T P L N Q L K G G T I V N V Y G V V K F F K P P Y L S K G I D Y C S V V T I V, D Q T N V K L T C L L F S G N Y E A L P I I Y K N G D I V R F H R L K I Q V Y K K E T Q G I T S S G, F A S L T F E G T L G A P I I P R T S S K Y E N F I T E D H K M V E A L R V W A S T H M S P S W I L, K L C D V Q P M Q Y F D L T C Q L L G K A E V D G A S F L L K V W D G T R T P F P S W R V L I Q D, L V L E G D L S H I H R L Q N L T I D I L V Y D N H V H V A R S L K V G S F L R I Y S L H T K L G S, M N S E N Q T M L S L E F H L H G G T S Y G R G I R V L P E S N S D V D Q L K K D L E S A N L T A, Q H S D V I Q Q S E P D D S F P S S G S V S L Y E V E R C Q O L S A T I L I D H O Y L E R T P L C A, I L L K Q K A P Q Q Y R I R A K L R S Y K P R R L F Q S V K L H C P K C H L Q E V P H E G D L D I I, F D D G A T K T P D V K L Q N T S L Y D S K I W T T K N O K G R K V A H F V K N N G I L P L S N E, C L L L I E G G T L S E I O K L S N K F N S V I P V R S G H E D L E L L D S A P F L I Q G T I H H, Y G C K Q C S S L R S I Q N L N S L V D K T S W I P S S V A E A L G I V P L Q V Y F V M T F T L D D, G T G V L E A Y L M D S D K F Q I P A S E V L M D D D L O K S V D M I M D M F C P P G I K I D A Y, and P W L E C F I K S Y N V T N G T D N Q I C Y Q I F D T I V A E D V I.

Ilustración 42: Proteína 3KJP mutada y secuencia de aminoácidos

En la siguiente Ilustración, se observa una región marcada en verde, que corresponde con las secuencias que coinciden de la proteína mutada con respecto a la proteína sin mutar. La flecha roja, indica que en la proteína consenso existe una región de la molécula que no se encuentra en la proteína mutada.

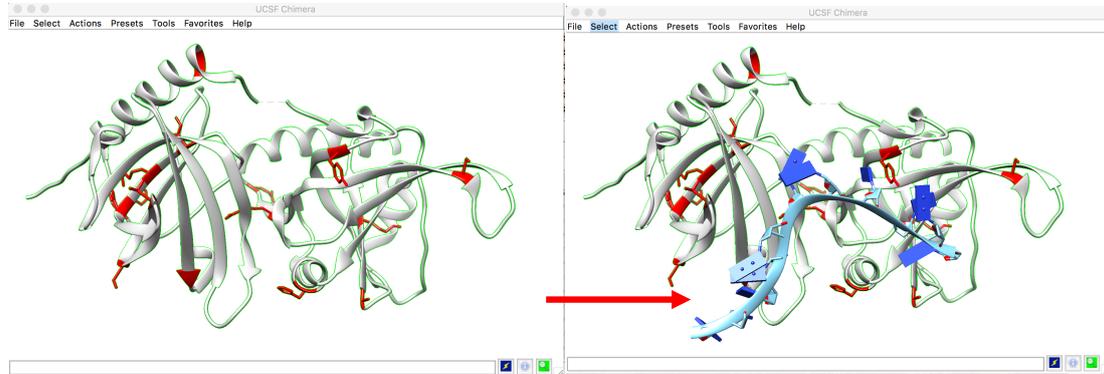


Ilustración 43: A la izquierda, la proteína mutada 3KJP. A la derecha, la proteína consenso 3KJP donde se muestra la región que no se ha codificado en la proteína mutada

En resumen, la mutación de la proteína 3KJP conlleva a plantearnos un estudio más exhaustivo en cuanto al funcionamiento de la misma y de las consecuencias que puede tener en la LLC, ya que las alteraciones en los telómeros también están asociadas al envejecimiento celular. Una mutación en este gen probablemente conlleve a que los linfocitos a envejecen prematuramente y conjuntamente con una mutación del gen BIRC3 que impide la apoptosis celular, se podría dar el caso de encontrarse con linfocitos B envejecidos que no reciben la señal de apoptosis.

Esta conclusión, es una teoría asociada a la consecuencia de la posible mutación de los dos genes de estudio a la vez.

## 6. Estrés Oxidativo

El estrés oxidativo es un proceso en el que interviene la oxidación. En bioquímica se considera oxidación a todo proceso en el que ocurre una pérdida de electrones, captación de oxígeno o una cesión de hidrógeno (deshidrogenación). Para compensar la oxidación, en el proceso de la reducción se captan electrones o se pierden oxígenos. Los procesos de oxidación van acompañados de procesos de reducción<sup>(32)</sup>.

El proceso oxidativo conlleva la liberación de radicales libres de oxígeno (RLO) que dañan a las moléculas, entre ellas, a los ácidos nucleicos, por lo que puede alterar un proceso celular. Cuando existe un exceso de radicales libres (RL) el equilibrio oxidación-reducción se ve alterado por lo que se produce el llamado estrés oxidativo.

Para contrarrestar los efectos del estrés oxidativo, se recomienda el consumo de antioxidantes, que son sustancias que proceden de alimentos como: Té verde, nueces, frutas, verduras, etc. Estas sustancias son capaces de oponerse a la acción del oxígeno, pero no sólo existen antioxidantes procedentes de la alimentación, también hay enzimas que son capaces de metabolizar los RL y otras sustancias como la vitamina C, vitamina E o ácido ascórbico, por ejemplo,<sup>(32)</sup> que también son utilizados como biomarcadores del estrés oxidativo.

El uso de biomarcadores es útil, debido a que los enfermos de cáncer presentan valores elevados de RL en el plasma. También hay que tener en cuenta que los enfermos de cáncer suelen presentar, además de las células neoplásicas (células de tipo tumoral), células inflamatorias, ya que los tratamientos oncológicos contribuyen a elevar los niveles del estrés oxidativo y pueden afectar a la correcta nutrición de los pacientes<sup>(33)</sup>.

Enfocando el estrés oxidativo a nivel genético, se realizó una búsqueda en la base de datos de NCBI y se eligió el gen OXSR1<sup>9</sup>, porque es un gen que se activa según el nivel de estrés ambiental.

## 6.1. Gen OXSR1 “Oxidative Stress Responsive 1”

El gen OXSR1 pertenece a la familia de proteínas Serina (Ser) / Treonina (Thr) quinasa, que se encarga de la regulación de las quinasa posteriores (hacia 3') en respuesta al estrés ambiental y puede también jugar un papel en la regulación del citoesqueleto de la actina<sup>(34)</sup>.

El gen se encuentra localizado en el cromosoma 3, concretamente en la región: 3p22.2 y se encuentra localizado entre las bases 38.165.051 a 38.255.488, según la anotación GRCh38.p7.

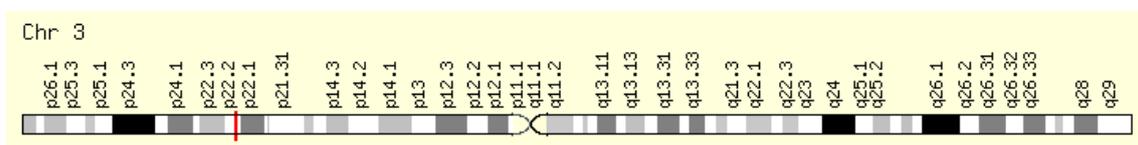


Ilustración 44: Localización del gen OXSR1 en el cromosoma 3

El número de acceso en GenBank es: NC\_000003.12, pero utilizaremos la secuencia del gen OXSR1 anotada en GenBank con el número de acceso: NM\_005109.2 que consta de 4.525 bp en mRNA (o ARNm). La elección de esta secuencia es porque es ARNm, por lo que es interesante conocer si sufre algún tipo de mutación que pueda ser asociada a la LLC.

<sup>9</sup> <https://www.ncbi.nlm.nih.gov/gene/9943>

Utilizando las herramientas que nos ofrece ICGC, se observa que el gen OXSR1 está presente en mutaciones asociadas a la LLC, como se muestra en la siguiente tabla:

ID	DNA change	Type	Consequences	# Donors affected	
				Total	Across all Projects
MU52436746	chr3:g.38213517A>G	single base substitution	Intron: <i>OXSR1</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU54847829	chr3:g.38237509C>A	single base substitution	Intron: <i>OXSR1</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU36717247	chr3:g.38220117T>A	single base substitution	Intron: <i>OXSR1</i>	1 / 2 (50.00%)	2 / 10,648 (0.02%)
MU45061632	chr3:g.38267646C>G	single base substitution	Downstream: <i>OXSR1</i> Intron: <i>OXSR1</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)
MU54225894	chr3:g.38285194G>T	single base substitution	Intron: <i>OXSR1</i>	1 / 1 (100.00%)	1 / 10,648 (0.01%)

Ilustración 45: La tabla informativa sobre el número de mutación y tipo de mutación

Actualmente, hay 5 mutaciones identificadas en el proyecto CLLE-ES, siendo éste el mismo proyecto que recogía las mutaciones de los genes BIRC3 y POT1.

Como se observa en la tabla, todas las mutaciones son por sustitución de una base en cadenas de intrones o en cadenas posteriores hacia 3'.

En la siguiente imagen, se muestra con detalle un ejemplo de sustitución de la mutación número MU52436746:



Ilustración 46: Resultado que muestra información sobre el tipo de mutación y la región de ADN afectada

Como se observa en el recuadro inferior, el nucleótido A ha mutado a G. Aunque la mutación es en este caso se produce en intrones, sí puede afectar a la cadena final de ARNm, ya que en los intrones hay secuencias consenso que son responsables de la correcta maduración del ARNm.

Si un cambio de nucleótido afecta a una secuencia, se podría dar que el proceso de corte y empalme (splicing) afecte al exón y este sea incorrecto, por lo que el ARNm maduro puede ser no funcional. <sup>(35)</sup>

En conclusión, aunque no existe una homología entre los genes BIRC3 y POT1 con el gen OXSR1, se podría considerar el estudio sobre este gen porque probablemente también podría ayudar a establecer un posible marcador para la LLC, o establecer una relación entre los demás genes relacionados con la LLC. Debido a que existe homología entre los genes BIRC3 y POT1 en el “Anexo I” estudiamos si este gen también tiene secuencias homólogas con retrovirus.

## 6.2. Estrés oxidativo, ARNm y la LLC

Las herramientas bioinformáticas nos ayudan a la búsqueda de secuencias homólogas entre cadenas de ADN o similitud en cadenas de proteínas. También pueden ayudar a construir árboles filogenéticos, encontrar ortólogos<sup>10</sup>, parólogos<sup>11</sup>, ancestros comunes y entender la evolución de especies, ya que muchas especies han evolucionado gracias a las mutaciones de sus genes.

Utilizar las herramientas bioinformáticas para la detección del estrés oxidativo es algo difícil, ya que el estrés oxidativo es un desequilibrio bioquímico que se produce a nivel plasmático donde se generan reacciones REDOX (reducción – oxidación) y la ausencia de algunas sustancias conllevan al desequilibrio y a la presencia de RL que pueden penetrar en la célula. En el caso del cáncer, sí podemos utilizar las herramientas bioinformáticas para la detección de las mutaciones genéticas.

En el cáncer, las mutaciones no son un proceso evolutivo que conlleve la mejora de la especie, sino que causa la no evolución de una parte de ella. Es importante comprender que muchas de las mutaciones actuales son debidas a mutaciones exteriores que afectan de alguna forma a la célula.

Unas de las vías actuales de investigación acerca de la causa de algunas enfermedades, entre ellas el cáncer, es el estrés oxidativo (explicado con anterioridad) y como éste puede dañar un gen de un cromosoma para que codifique mal una proteína. Los factores externos influyen en el estrés oxidativo ya que se produce un desequilibrio de los radicales libres con el proceso de reducción de los mismos.

---

<sup>10</sup> Los genes ortólogos, son secuencias homólogas que se han separado cuando una especie diverge en dos especies separadas, es decir, divergen del mismo gen.

<sup>11</sup> Dos o más secuencias homólogas se dicen que son parálogas si un gen de un organismo se duplica para ocupar dos posiciones diferentes dentro de un mismo genoma.

Los radicales libres pueden afectar a cualquier parte de la célula y por consiguiente a cualquier cromosoma, pero parece ser que hay algunos genes que son más susceptibles a una mutación causada por el estrés oxidativo que otros genes.

Si este tipo de radicales libres afecta a una secuencia de ADN, la cadena de ARNm también será afectada y por tanto el resultado de la síntesis de ARNm será una proteína que sea defectuosa y conlleve a desarrollar un tipo de cáncer.

En el caso de la LLC, el estrés oxidativo puede afectar a las células madre encargadas de generar linfocitos B haciendo que estos sean defectuosos y sean de tipo cancerígeno. Se puede dar el caso que los RL afecten a la secuencia de ADN que codifica a los linfocitos B, por ejemplo, cuando un linfocito B se clona, en el proceso interviene el microARN, que es una cadena corta de ARNm, que puede ser afectada por un RL en el momento que sale del núcleo celular, por lo que el clon sería defectuoso.

Si este clon de linfocito B se multiplica, también será defectuoso y si no se activa el sistema de control de apoptosis celular, estas células seguirán clonándose generando linfocitos cancerígenos.

En conclusión, las investigaciones muestran una relación entre enfermedades de tipo cáncer, concretamente de la LLC, y el estrés oxidativo. Es importante el uso de las herramientas bioinformáticas, de esta forma, se puede entender cómo se produce la mutación de una célula que no está afectada por estrés oxidativo comparándola con otra célula que sí esté afectada por estrés oxidativo y así, comprender qué alteraciones se producen a nivel de ADN o a nivel de proteína.

## 7. Retrovirus y la LLC

En el proceso de búsqueda de homologías entre posibles genes causantes de la LLC, se han encontrado homologías entre algunos retrovirus y la LLC.

Es conocido que entre un 5-8% del genoma humano está formado por cadenas de retrovirus endógenos (ERV)<sup>(36)</sup>. La palabra endógeno, significa que se origina o nace en el interior y aplicándolo a nivel biológico, lo podemos definir como la célula que se forma dentro de otra. Los retrovirus endógenos son elementos virales endógenos en el genoma, que son similares entre si y pueden ser derivado de un retrovirus.

Un retrovirus es un virus que pertenece a la familia de Retroviridae y consta de un genoma de ARN monocatenario y replicándose de manera inusual a través de una forma intermedia de ADN bicatenario. Este proceso es llevado a cabo por

una encima que se llama retrotranscriptasa o transcriptasa inversa, que está presente en los retrovirus. La función de la transcriptasa inversa es la de sintetizar ADN de doble cadena utilizando como molde un ARN monocatenario. Una vez que se ha realizado el proceso de ARN monocatenario a ADN, éste se inserta dentro del ADN propio de la célula infectada donde se comporta como un gen más.

Los retrovirus son causantes de muchas enfermedades, incluyendo cánceres.

## 7.1 Homologías entre el gen BIRC3 y retrovirus

La búsqueda de homologías entre el gen BIRC3 y los retrovirus se realizó de una forma sencilla. Por una parte, tenemos la secuencia del gen BIRC3 y por otra parte realizamos las consultas en las bases de datos que proporciona la herramienta BLASTN alojada en el NCBI.

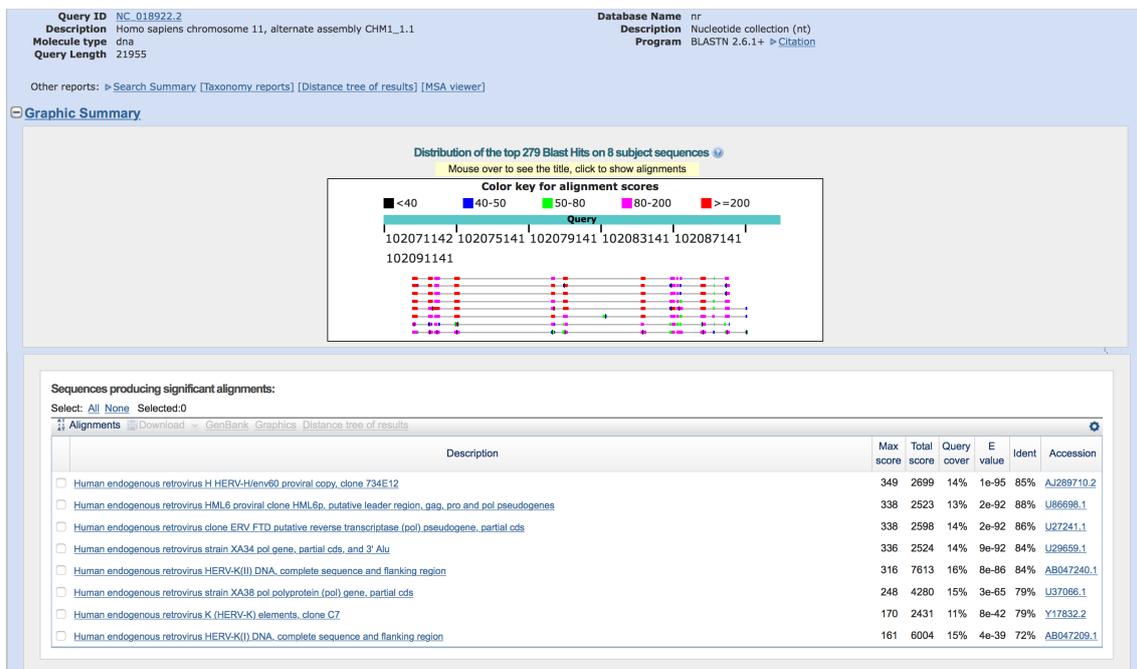


Ilustración 47: Resultado de la búsqueda BLASTN del gen BIRC3 y retrovirus HERV

Cómo se observa en la imagen, existe homología entre *Human endogenous retrovirus* (HERV) del tipo H y del tipo K con la secuencia del gen BIRC3, obteniéndose como mejor resultado un *Identity* del 85%, con un *Cover* del 14% y un *E-value* 1e-95. Esto indica que hay partes conservativas en la secuencia de BIRC3 que son homólogas a con HERV, es decir, que esas partes conservativas tienen un ancestro en común, en concreto el retrovirus endógeno humano, que durante la evolución ha formado parte de la secuencia del ADN humano. La secuencia completa puede ser consultada aquí: <https://goo.gl/6LcAYi>

Si utilizamos la secuencia del gen POT1 y la comparamos con el HERV, se obtienen los siguientes resultados:

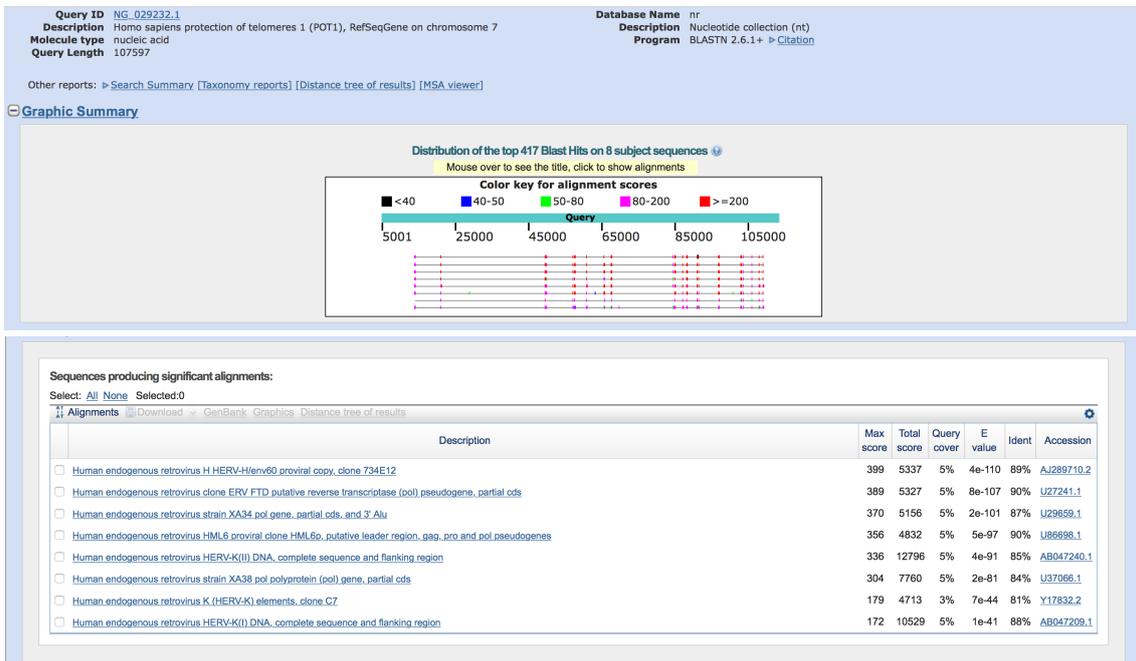


Ilustración 48: Resultado de la búsqueda BLASTN del gen POT1 y retrovirus HERV

Como se observa en la imagen, los resultados son prácticamente iguales a los del gen BIRC3, aunque si nos fijamos en la columna del Cover, éste valor obtiene un máximo del 5%, inferior al de BIRC3, esto es debido a que la secuencia del gen POT1 es mayor (107.597 bp) que la del gen BIRC3 (21.955 bp). La secuencia completa se puede consultar aquí: <https://goo.gl/uUe01t>

En los resultados obtenidos en ambos genes, se observa que existe una codificación parcial Alu (en la tercera línea del resultado de alineamientos). Una secuencia Alu es una secuencia de ADN de aproximadamente 300 bp y se encuentra en muchos lugares del genoma de los primates. Estas secuencias se analizan mediante técnicas de PCR, que consiste en la reacción en la cadena de polimerasa, técnica muy empleada para obtener copias de fragmentos de ADN.

Las secuencias Alu son muy abundantes dentro del genoma humano y estas secuencias se sitúan hace aproximadamente 65 millones de años, que coincide con el origen de los primates. Es importante este dato, ya que las inserciones con estas secuencias han sido vinculadas con enfermedades heredables en humanos, incluyendo varias formas de cáncer. <sup>(36)</sup>

Podemos ampliar la búsqueda con otro tipo de retrovirus, entre ellos el VIH-1 que se comenta a continuación.

## 7.2. Retrovirus HTLV-I/II y HTLV-III (VIH-1)

El Virus de la Inmunodeficiencia Humana tipo I o VIH-1, es un lentivirus que pertenece a la familia de los retrovirus y causa con el tiempo el “Síndrome de la Inmunodeficiencia Adquirida” (SIDA),<sup>(37)</sup> ya que debilita los linfocitos T CD4+.<sup>(38)</sup>

En concreto, uno de los nombres que se dio al VIH-1 en sus inicios fue HTLV-III, *Human T-Lymphotropic Virus type III*<sup>(39)</sup>, lo que viene a significar: Virus Linfotrópico de células T Humanas tipo III.

Existen además dos tipos de HTLV: el HTLV-I que es un retrovirus de tipo oncológico que puede provocar leucemia o linfoma de células T en el adulto y el HTLV-II, que no se conoce todavía qué tipo de enfermedad puede llegar a desarrollar, aunque no se encuentra relación directa con la leucemia.<sup>(40)</sup>

Cuando se realizaron las búsquedas de homologías de los genes BIRC3 y POT1, se obtuvieron como resultados secuencias asociadas al VIH-1, no se obtuvo como resultado ningún HTLV-I/II. Debido a esto, se compararon los virus VIH-1 con HTLV-I/II obteniéndose, como se puede ver a continuación, que el virus VIH-1 no tiene homología aparentemente con HTLV-I/II, aunque pertenezcan a la misma familia. Este es el motivo por el cual sólo se estudiará la homología de retrovirus entre VIH-1 y los genes BIRC3 y POT1.

Se ha tomado por un lado el virus HTLV-I con número de acceso en GenBank: J02029.1 y se ha realizado una consulta con la herramienta BLASTN en NCBI, en la base de datos nr/nt. Se han introduciendo las siguientes nomenclaturas en el campo *Organism* (campo donde se introduce el nombre del organismo a buscar o taxid, que es un número asociado al organismo):

- HIV (taxid:11676)
- HIV-1 (taxid:11676)
- human immunodeficiency virus 1 HIV-1 (taxid:11676)
- AIDS virus (taxid:12721)
- AIDS virus (taxid:11709)
- Aids-associated retrovirus (taxid:11966)
- Simian-Human immunodeficiency virus (taxid:57667)
- Simian-Human immunodeficiency virus 1 (taxid:1563316)

El motivo por el cual se introducen tantos nombres, es para ampliar el campo de búsqueda, ya que el virus VIH-1 (HIV-1 en inglés), puede ser accedido por múltiples nombres.

El resultado de la búsqueda es mostrado en la siguiente imagen:

**Query ID** J02029.1  
**Description** Human T-cell lymphotropic virus type 1 (proviral), complete genome  
**Molecule type** nucleic acid  
**Query Length** 9058

**Database Name** nr  
**Description** Nucleotide collection (nt)  
**Program** BLASTN 2.6.1+ > Citation

Other reports: > Search Summary [Taxonomy reports] [Distance tree of results] [MSA viewer]

**Graphic Summary**

Distribution of the top 5 Blast Hits on 5 subject sequences  
 Mouse over to see the title, click to show alignments

Color key for alignment scores  
 ■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200

Sequences producing significant alignments:

Select: All None Selected:0

Alignments	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	HIV-1 isolate 030-00777-8_03OCT96 from USA pol protein (pol) gene, partial cds	50.0	50.0	0%	0.005	91%	KF729844.1
<input type="checkbox"/>	HIV-1 isolate 030-00186-4_31OCT96 from USA pol protein (pol) gene, partial cds	50.0	50.0	0%	0.005	91%	KF729839.1
<input type="checkbox"/>	HIV-1 isolate 10164741 pol protein (pol) gene, partial cds	50.0	50.0	0%	0.005	91%	HQ684963.1
<input type="checkbox"/>	HIV-1 isolate NC6002-2000 from USA pol polyprotein (pol) gene, partial cds	50.0	50.0	0%	0.005	91%	AY032245.1
<input type="checkbox"/>	HIV-1 isolate SP148_00 from Brazil protease (pol) gene, partial cds	50.0	50.0	0%	0.005	91%	AY441845.1

**Download** GenBank Graphics

HIV-1 isolate 030-00777-8\_03OCT96 from USA pol protein (pol) gene, partial cds  
 Sequence ID: KF729844.1 Length: 1302 Number of Matches: 1

Range 1: 61 to 95 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
50.0 bits(54)	0.005	32/35(91%)	0/35(0%)	Plus/Plus

Query 2251 GAAGCTTTACTAGATACAGGACAGACATGACAGT 2285  
 Sbjct 61 GAAGCTTTACTAGATACAGGACAGACATGACAGT 95

**Download** GenBank Graphics

HIV-1 isolate 030-00186-4\_31OCT96 from USA pol protein (pol) gene, partial cds  
 Sequence ID: KF729839.1 Length: 1302 Number of Matches: 1

Range 1: 61 to 95 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
50.0 bits(54)	0.005	32/35(91%)	0/35(0%)	Plus/Plus

Query 2251 GAAGCTTTACTAGATACAGGACAGACATGACAGT 2285  
 Sbjct 61 GAAGCTTTACTAGATACAGGACAGACATGACAGT 95

**Download** GenBank Graphics

HIV-1 isolate 10164741 pol protein (pol) gene, partial cds  
 Sequence ID: HQ684963.1 Length: 1968 Number of Matches: 1

Range 1: 52 to 86 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
50.0 bits(54)	0.005	32/35(91%)	0/35(0%)	Plus/Plus

Query 2251 GAAGCTTTACTAGATACAGGACAGACATGACAGT 2285  
 Sbjct 52 GAAGCTTTACTAGATACAGGACAGACATGACAGT 86

Ilustración 49: Resultado de la búsqueda con BLASTN del virus HTLV-I y HIV-1

En conclusión, no se puede decir que exista una homología, ya que sólo hay una secuencia de 32 bp con un *E-value* de 0.005, y teniendo en cuenta que hay miles de secuencias de VIH-1 en las bases de datos nr/nt, estos resultados no son concluyentes de una homología y menos cuando se trata de un virus que se supone que es de la misma familia de retrovirus.

Si hacemos la misma búsqueda con el virus HTLV-II cuyo número de acceso en GenBank es AF412314.1, no obtenemos ningún resultado positivo de homología:

**Query ID** AF412314.1  
**Description** Human T-cell lymphotropic virus type 2, complete genome  
**Molecule type** nucleic acid  
**Query Length** 8962

**Database Name** nr  
**Description** Nucleotide collection (nt)  
**Program** BLASTN 2.6.1+ > Citation

No significant similarity found. For reasons why, click here

Other reports: > Search Summary

Ilustración 50: Resultado de la búsqueda con BLASTN del virus HTLV-II y HIV-1

Por lo que se deduce que el HTLV I y II no son de la misma familia que el HTLV-III o VIH-1. Se podría buscar algún tipo de ancestro común o similitud utilizando herramientas como TBLASTX, que convierte la secuencia de ADN en secuencias de aminoácidos y después las compara, pero no es el ámbito de este estudio, por lo que se deja una puerta abierta para una futura investigación.

### 7.2.1 Homologías entre VIH-1 y genes BIRC3 y POT1

Como se ha comentado en el apartado anterior, se procede al estudio de la homología de los genes BIRC3 y POT1 con el VIH-1.

Por un lado, tenemos la secuencia de BIRC3 y por otro lado usaremos todos los accesos disponibles al VIH-1 en la base de datos nr/nt utilizando la herramienta BLASTN en NCBI.

En la siguiente imagen se muestran los resultados limitado a 100 alineamientos.

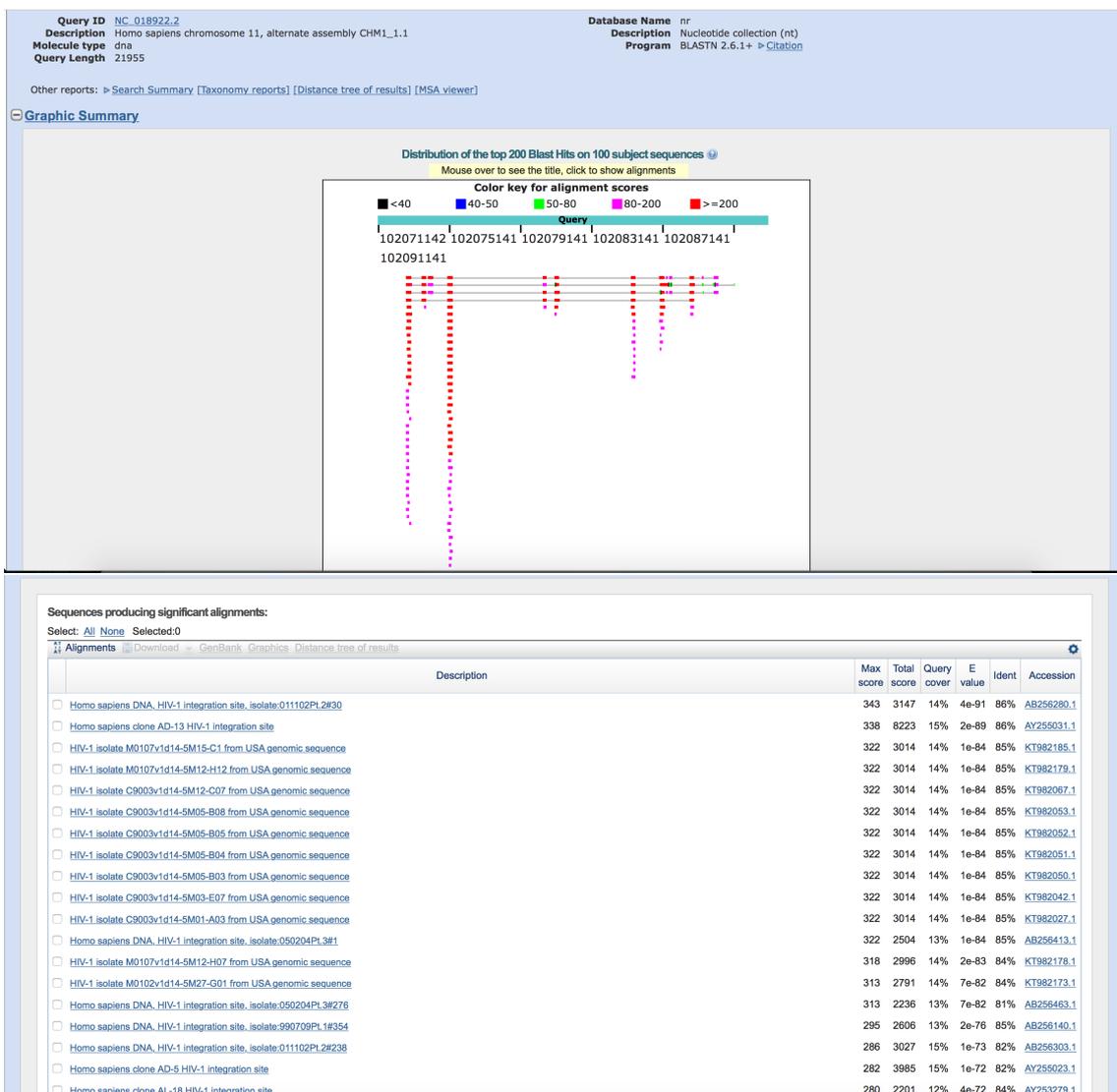


Ilustración 51: Resultado de la búsqueda con BLASTN del gen BIRC3 y HIV-1

Como se puede observar, existe una homología entre el gen BIRC3 y el VIH-1, en concreto en la parte llamada integración LTR o integration site. Los valores obtenidos son del 86% de identidad y un *E-value* de 4e-91. Estos valores son parecidos a los obtenidos por el virus HERV, de hecho, si se observa el alineamiento completo que se puede consultar aquí: <https://goo.gl/5OYj31>, se observan que las secuencias son muy parecidas.

Si realizamos la misma búsqueda para el gen POT1, se obtienen los siguientes resultados que se muestran en la siguiente imagen:

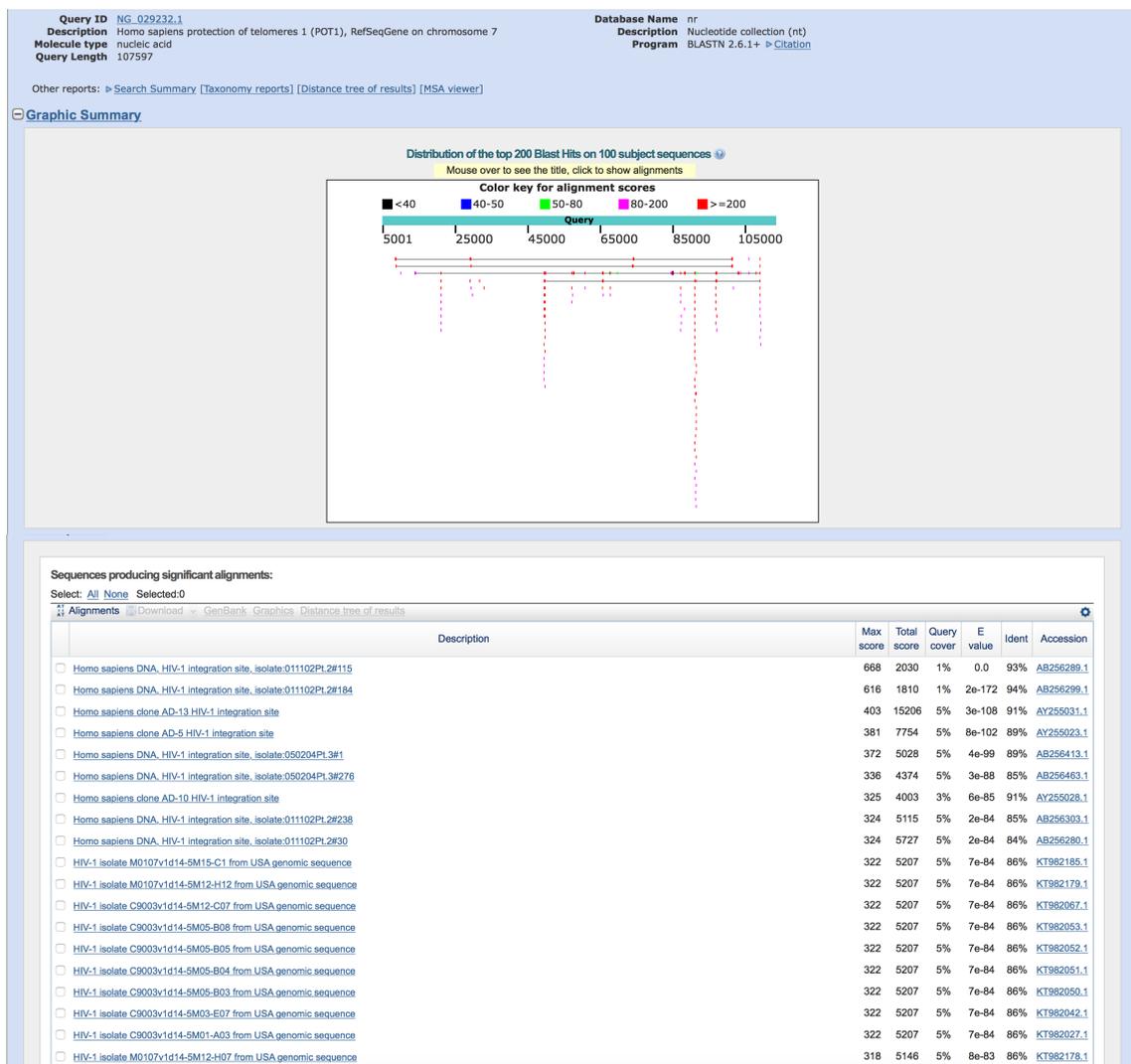


Ilustración 52: Resultado de la búsqueda con BLASTN del gen POT1 y HIV-1

Como se puede observar, existe homología entre el gen POT1 y el virus VIH-1 y se obtienen unos valores mejores con respecto al anterior gen BIRC3. El resultado del alineamiento se puede consultar aquí: <https://goo.gl/BmDfdT>

En conclusión, se puede deducir que existe homología entre los dos genes de estudio y la parte integración del virus VIH-1.

## 8. LLC y el gen TP53

Una de las causas por las que la LLC tiene un complicado pronóstico de curación es debido a un mal funcionamiento del gen TP53. Este gen codifica la proteína tumoral p53 (o p53) que actúa como supresor de tumores, lo que significa que esta proteína es la encargada de regular la división celular, vigilando que las células no tengan un crecimiento o una división demasiado rápida o de manera incontrolada.

La proteína p53 se encuentra en el núcleo de las células de todo el cuerpo y se une directamente al ADN. Cuando la célula es dañada por agentes externos como rayos ultravioletas, radiación, sustancias químicas, etc., la p53 se encarga de determinar si el ADN será reparado o si la célula dañada se autodestruirá (apoptosis).

En el caso de que el ADN pueda ser reparado, la p53 es la encargada de la activación de otros genes que participan en la reparación del daño. Si ocurre el caso contrario, es decir, si el ADN no puede ser reparado, la p53 impide que la célula se divida, facilitando que se produzca la apoptosis<sup>(38)</sup>. Llegados a este punto, se ha de tener presente que el gen BIRC3 relacionado con la LLC se encarga de la apoptosis celular, por lo que, si este gen no ejerce correctamente su función, no podrá programar la muerte celular.

La localización del gen TP53, se encuentra en el cromosoma 17, concretamente: 17p13.1, entre los pares de bases: 7.668.402 hasta 7.687.550, según la anotación GRCh38.p7.

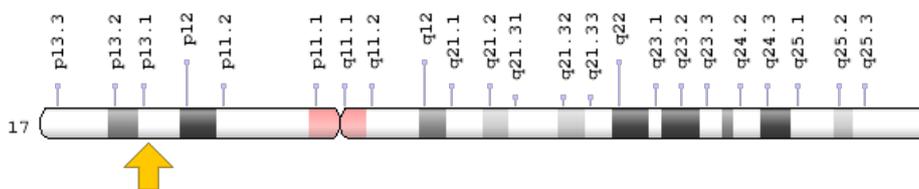


Ilustración 53: Localización del gen TP53 en el cromosoma 7

Para ver el comportamiento del gen TP53 en la LLC, se consultan los datos de mutaciones que ofrece ICGC. Realizando una búsqueda en el servidor de datos, aparece que sí existe mutaciones de este gen asociadas a pacientes con la LLC, bajo el nombre de proyecto CLLE-ES. Como se observa en la siguiente imagen,

la tabla muestra las edades de los pacientes, que están comprendidas entre los 60 y 80 años, lo que nos indica que concuerda con los datos en los que la LLC es detectada:

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types:											# Mutations	# Genes
							SSM	CNSM	SISM	SGV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN		
DO223401	CLLE-ES	Blood	Male	62	A	3,364	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO52706	CLLE-ES	Blood	Female	75	A	1,914	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO7166	CLLE-ES	Blood	Male	62	A	1,890	✓	--	--	--	--	--	✓	✓	--	--	--	2	1
DO52720	CLLE-ES	Blood	Female	61	A	3,034	✓	✓	✓	--	--	--	✓	✓	--	--	--	1	1
DO52716	CLLE-ES	Blood	Female	61	A	1,310	✓	✓	--	--	--	--	✓	✓	--	--	--	1	1
DO6434	CLLE-ES	Blood	Female	45	A	3,407	✓	✓	✓	--	✓	--	✓	✓	--	--	--	1	1
DO52710	CLLE-ES	Blood	Male	86	A	1,177	✓	✓	--	--	--	--	✓	✓	--	--	--	1	1
DO6730	CLLE-ES	Blood	Male	67	A	2,604	✓	✓	✓	--	✓	--	✓	✓	--	--	--	2	1

Ilustración 54: Tabla que muestra el estudio de pacientes con mutación en el gen TP53 y su asociación con la LLC

En la siguiente tabla, se observa la información de las mutaciones relacionadas con el gen TP53 y la LLC.

ID	DNA change	Type	Consequences	# Donors affected	
				Total	Across all Projects
MU4807	chr17:g.7578190T>C	single base substitution	Missense: <i>TP53</i> Y127C, Y88C, Y220C Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i> Intron: <i>TP53</i>	2 / 58 (3.45%)	58 / 11,704 (0.50%)
MU25915	chr17:g.7576851A>C	single base substitution	Splice Donor: <i>TP53</i> Downstream: <i>TP53</i> Intron: <i>TP53</i>	1 / 1 (100.00%)	1 / 11,704 (0.01%)
MU1101758	chr17:g.7577566T>C	single base substitution	Missense: <i>TP53</i> N239D, N146D, N107D Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i>	1 / 5 (20.00%)	5 / 11,704 (0.04%)
MU3103824	chr17:g.7576851A>G	single base substitution	Splice Donor: <i>TP53</i> Downstream: <i>TP53</i> Intron: <i>TP53</i>	1 / 3 (33.33%)	3 / 11,704 (0.03%)
MU28655667	chr17:g.7577545T>A	single base substitution	Missense: <i>TP53</i> M153L, M246L, M114L Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i>	1 / 2 (50.00%)	2 / 11,704 (0.02%)
MU11229	chr17:g.7577547C>T	single base substitution	Missense: <i>TP53</i> G113D, G152D, G245D Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i>	1 / 18 (5.56%)	18 / 11,704 (0.15%)
MU70725	chr17:g.7578403C>T	single base substitution	Missense: <i>TP53</i> C44Y, C176Y, C83Y Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i>	1 / 19 (5.26%)	19 / 11,704 (0.16%)
MU54541660	chr17:g.7577953C>T	single base substitution	Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i> Intron: <i>TP53</i>	1 / 1 (100.00%)	1 / 11,704 (0.01%)
MU25506	chr17:g.7578413C>G	single base substitution	Missense: <i>TP53</i> V173L, V41L, V80L Upstream: <i>TP53</i> Exon: <i>TP53</i> Downstream: <i>TP53</i>	1 / 5 (20.00%)	5 / 11,704 (0.04%)

Ilustración 55: La tabla informativa sobre el número de mutación y tipo de mutación del gen TP53

Como se observa en la tabla, las mutaciones se deben a un cambio de sustitución de base y en la mayoría de los casos es en exones y en las cadenas posteriores hacia 3' y anteriores hacia 5'.

## 8.1. Homología entre el gen TP53 y los genes BIRC3 y POT1

A continuación, se estudiará la homología del gen TP53 con los genes BIRC3 y POT1 por si existiera una relación entre ellos que pueda asociar algún patrón de comportamiento. Para ello utilizaremos la herramienta BLASTN en el portal NCBI.

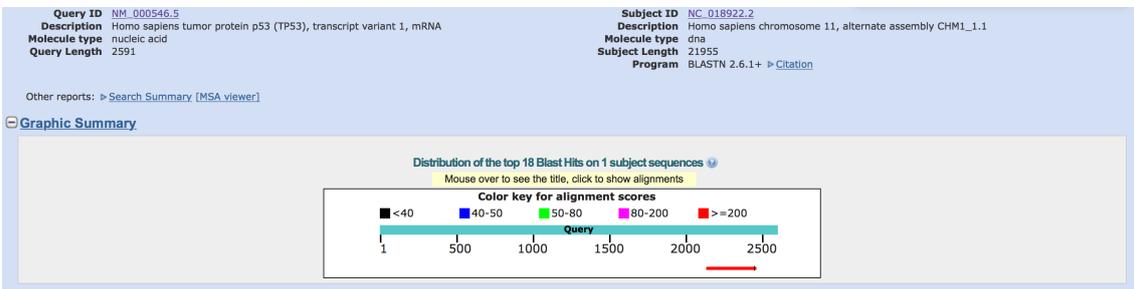
En la siguiente imagen, se observan los resultados de la homología entre el gen TP53 y POT1:



Ilustración 56: Resultado de BLASTN del gen TP53 y POT1

Como resultado, se obtiene que el gen POT1 tiene una homología con respecto a TP53 con un valor de identidad del 79% y un *E-Value* de  $9e-64$ . El alineamiento completo puede ser consultado aquí: <https://goo.gl/x2uADb>. Se aprecia que es al final de la secuencia del gen TP53 en donde se encuentra la homología con respecto a POT1.

Por último, en la siguiente imagen se muestra la homología entre el gen p53 y BIRC3:



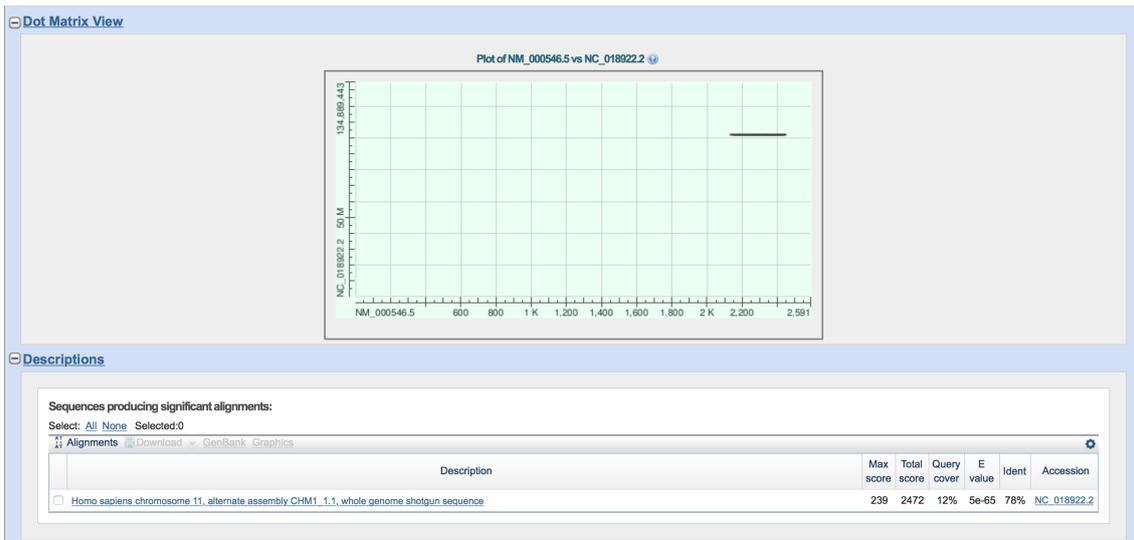


Ilustración 57: Resultado de BLASTN del gen TP53 y BIRC3

Como podemos observar en el resultado de la alineación, existe una homología al final del gen TP53 con un valor de identidad del 72% y un valor *E-value* de 5e-65. El alineamiento completo se puede consultar aquí: <https://goo.gl/3Qowf1>. Como se observa, el alineamiento es muy similar al POT1.

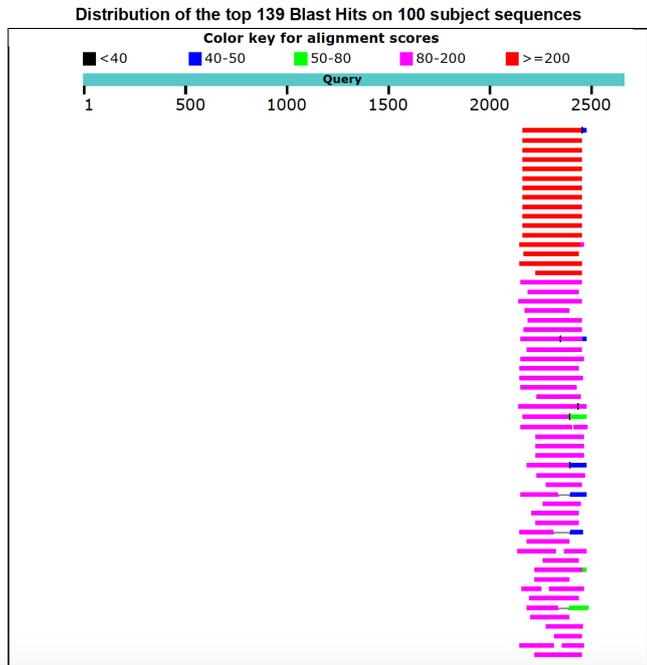
Esto resultado lleva a pensar si también existirá una homología entre el gen TP53 y el virus VIH-1, que se tratará en el siguiente apartado:

## 8.2. Gen TP53 y VIH-1

Como se observó en apartados anteriores, existe una homología entre los genes BIRC3 y POT1 con el virus VIH-1. Realizando las mismas búsquedas que con los anteriores genes y utilizando la herramienta BLASTN en el portal NCBI, obtenemos los siguientes resultados:

**Query ID** [NM\\_000546.5](#)  
**Description** Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA  
**Molecule type** nucleic acid  
**Query Length** 2591  
**Database Name** nr  
**Description** Nucleotide collection (nt)  
**Program** BLASTN 2.6.1+

### Graphic Summary



### Descriptions

Sequences producing significant alignments:

Description	Max score	Total score	Query cover	E value	Ident	Accession
Homo sapiens DNA, HIV-1 integration site, isolate:011102Pt.2#30	257	305	11%	6e-66	80%	<a href="#">AB256280.1</a>
HIV-1 isolate M0107v1d14-5M15-C1 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982185.1</a>
HIV-1 isolate M0107v1d14-5M12-H12 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982179.1</a>
HIV-1 isolate M0107v1d14-5M12-H07 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982178.1</a>
HIV-1 isolate C9003v1d14-5M12-C07 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982067.1</a>
HIV-1 isolate C9003v1d14-5M05-B08 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982053.1</a>
HIV-1 isolate C9003v1d14-5M05-B05 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982052.1</a>
HIV-1 isolate C9003v1d14-5M05-B04 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982051.1</a>
HIV-1 isolate C9003v1d14-5M05-B03 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982050.1</a>
HIV-1 isolate C9003v1d14-5M03-E07 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982042.1</a>
HIV-1 isolate C9003v1d14-5M01-A03 from USA genomic sequence	250	250	10%	8e-64	80%	<a href="#">KT982027.1</a>
HIV-1 isolate M0102v1d14-5M27-G01 from USA genomic sequence	232	232	10%	2e-58	79%	<a href="#">KT982173.1</a>
Homo sapiens clone AD-13 HIV-1 integration site	212	615	11%	2e-52	77%	<a href="#">AY255031.1</a>

Ilustración 58: Resultado de la búsqueda en BLASTN del gen TP53 y VIH-1

Como se observa en los resultados, existen al menos más de 100 alineaciones que devuelven una homología entre el gen TP53 y el virus VIH-1 en la región de integración “Integration site”. Los valores que nos devuelve BLASTN en cuanto a la identidad son del 80% con un *E-Value* de 6e-66. También se encuentra esta homología al final de la secuencia del gen TP53. Los resultados del alineamiento se pueden consultar aquí: <https://goo.gl/Dnvtck>

En resumen, según vemos en estos resultados, sabemos que el virus VIH-1 es un lentivirus de origen primate (el virus de la inmunodeficiencia simia “saltó” de los simios a los humanos) y que ataca a las células del sistema inmunitario, por lo que parte de su código genético coincide con parte del código genético de algunos genes relacionados con la LLC. El motivo de esta coincidencia debería de ser estudiado con más detenimiento.

## Capítulo 3: Conclusiones

### 1. Conclusión final

Hoy en día, las herramientas de software para secuenciación de genomas y predicción genética han avanzado a gran escala. Este tipo de herramientas bioinformáticas son de gran utilidad para el estudio de los genes y el estudio de las mutaciones, lo que conlleva a una mejora de la prevención de cierto tipo de enfermedades.

Es muy importante mantener actualizadas las bases de datos genéticas con secuencias de genes de pacientes enfermos de cáncer, y así poder observar qué partes del gen han mutado y son causantes del desarrollo de una enfermedad.

En la elaboración de este trabajo se han llegado a las siguientes conclusiones:

- Los marcadores genéticos ayudan a la prevención temprana de la LLC, así como la predisposición genética de una persona a poder desarrollar la LLC.
- La búsqueda de homólogos a los genes BIRC3 y POT1, podría ayudar a ampliar la detección de la LLC.
- Se han encontrado homologías entre los genes: BIRC3, XIAP, GRIA3 y DOCK7.
- Se han encontrado homologías entre los genes: POT1, SCP2 y PPEF1.
- Las mutaciones en un gen causadas por delección, inserción o por sustitución, pueden alterar la codificación y el comportamiento de una proteína.
- Con respecto a las mutaciones, se ha mostrado información acerca de secuencias de dos genes: BIRC3 y POT1. Estos genes presentan secuencias mutadas en enfermos de LLC, según el portal ICGC.
- Se ha estudiado, por un lado, las regiones donde aparecen las mutaciones y, por otro lado, a través del portal "Cancer3D.org", la simulación de la molécula a nivel estructural, entendiendo de esa forma cómo afectan las mutaciones a la codificación de la proteína final, utilizando para tal estudio la herramienta "Chimera".
- El gen OXSR1 (asociado al control de estrés oxidativo), cuando sufre una mutación, está presente en pacientes con LLC,
- La búsqueda masiva de genes homólogos a BIRC3 y POT1, ha dado como resultado secuencias homólogas con retrovirus: HERV y VIH-1.
- El gen TP53, juega un papel muy importante en el desarrollo de enfermedades relacionadas con el cáncer, como se ha estudiado, una mutación de este gen también está relacionada con la LLC.

- Existe homología entre el gen TP53 con respecto a los genes BIRC3 y POT1, y al mismo tiempo estos tres genes contienen secuencias homólogas al VIH-1
- Sorprendentemente, no se encuentra homología en tres virus de la misma familia HTLV-I y HTLV-II con el HTLV-III (VIH-1) usando la aplicación BLASTN.

## 1.1 Discusión

La búsqueda de homólogos es una herramienta de predicción útil a la hora del diagnóstico de enfermedades hereditarias o relacionadas con el cáncer.

Una inversión en la secuenciación de pacientes con enfermedades relacionadas con el cáncer, ayudaría a comprender mucho mejor el curso de las mismas, así como en la prevención. Si tenemos secuencias de pacientes con genes asociados a un tipo de cáncer o a la LLC, se podrá ver qué regiones han mutado y se podrá realizar un diagnóstico mediante marcadores genéticos.

Al mismo tiempo, este tipo de información es válida para poder desarrollar tratamientos para combatir enfermedades. En el caso de la LLC, los tratamientos actuales tienen bastante éxito, pero cuanto antes sea el diagnóstico menos será el tiempo de tratamiento quimioterapéutico recibirá un paciente.

Por último, la relación existente entre partes de secuencias de genes relacionados con la LLC y el virus de la inmunodeficiencia humana VIH-1, ha sido un motivo sorprendente, lo que abre una posible vía de investigación futura.

## 2. Reflexión

Hubiera sido muy interesante poder estudiar los fármacos empleados para los tratamientos de la LLC, estudiar la secuencia de aminoácidos de estos tratamientos, su comportamiento a nivel molecular, el mecanismo de acción, etc.

En cuanto al estrés oxidativo, he intentado dar otro enfoque a nivel genético, ya que la mayoría de la bibliografía se basa en la influencia de los factores externos en cuanto a las reacciones REDOX y el déficit de antioxidantes, todo ello enfocado a un nivel nutricional.

## 3. Análisis crítico

El seguimiento de la planificación, como se ha comentado anteriormente, ya sido bastante difícil a causa de factores externos que han impedido que durante unas

semanas no se pudiera seguir lo planificado, pero dada las circunstancias, se ha conseguido cubrir la mayoría de objetivos marcados.

#### 4. Líneas de trabajo futuras

Como línea de trabajo futuro, queda pendiente buscar más genes relacionados con el estrés oxidativo y la LLC. Así como el estudio más exhaustivo de la relación de genes asociados al cáncer y retrovirus.

El estudio de la homología del retrovirus HERV y de genes asociados a la LLC podría ayudar a comprender la integración de los retrovirus en las secuencias de ADN humano y si esta integración podría afectar o no a la mutación de un gen.

Habría que ampliar el estudio en cuanto a la mutación de genes relacionados con el estrés oxidativo, en este trabajo se ha hablado del gen OXSR1 y como se ha encontrado mutado en pacientes con LLC.

Por último, investigar sobre la relación entre las secuencias de genes asociados con el cáncer y el VIH-1, ya que se han evidenciado partes homólogas ya que hay que intentar descartar que se traten de contaminaciones.<sup>12</sup>

---

<sup>12</sup> Contamination of genomic databases by HIV-1 and its possible consequences. A study in Bioinformatics."2014 - Miguel Romero Fernández-Bravo. <http://hdl.handle.net/10609/31361>

## Glosario

- ADN: Ácido Desoxirribonucleico.
- ARN: Ácido Ribonucleico.
- ARNm: Ácido Ribonucleico mensajero.
- Bp: Pares de bases.
- BLAST: Basic Local Alignment Search Tools.
- ERV: Endogenous RetroVirus.
- HTLV: Virus linfotrópico de células T Humanas.
- LLC: Leucemia Linfática Crónica.
- LTR: Long Terminal Repeat
- NCBI: National Center for Biotechnology Information.
- NFkB: Factor Nuclear potenciador de las cadenas ligeras Kappa de las células B activadas.
- REDOX: Reducción-Oxidación de una reacción química de electrones.

## Bibliografía

1. American Cancer Society. [En línea] <https://www.cancer.org/es/cancer/leucemia-linfocitica-cronica/acerca/que-es-leucemia-linfocitica-cronica.html>.
2. Hematología, Revista de. *medigraphic.com*. [En línea] <http://www.medigraphic.com/pdfs/hematologia/re-2014/re143d.pdf>.
3. Colls., José A. García Marco y. *Guías de consenso nacionales para el manejo de pacientes con leucemia linfocítica crónica*. s.l. : inScienceCommunications. pág. 6.
4. Devlin, Tomas M. *Bioquímica: libro de texto con aplicaciones clínicas*. s.l. : Editorial Reverté.
5. Déu, Hospital Sant Joan de. [www.guiametabolica.org](http://www.guiametabolica.org). [En línea] <https://www.guiametabolica.org/noticia/gen-proteina-0>.
6. Martínez, José Peña. Inmunología en Línea. [En línea] <http://www.inmunologiaenlinea.es/index.php/03mmm-9/b>.
7. Colls., Xose S. Puente et. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. [En línea] 05 de Jun de 2011. doi:10.1038/nature10113. <http://www.nature.com/nature/journal/v475/n7354/full/nature10113.html>.
8. NIH. NIH U.S. National Library of Medicine. [En línea] <https://ghr.nlm.nih.gov/primer/howgeneswork/genelocation>.
9. —. U.S. Department of Health & Human Services. [En línea] <https://ghr.nlm.nih.gov/gene/BIRC3#normalfunction>.
10. Herishanu, et al. American Society of Hematology. *The lymph node microenvironment promotes B-cell receptor signaling, NF-κB activation, and tumor proliferation in chronic lymphocytic leukemia*. [En línea] 2011. <http://www.bloodjournal.org/content/117/2/563.long?sso-checked=true>.
11. Gilmore, T.D. Introduction to NF-κB: players, pathways, perspectives». *Oncogene* 25 (51): 6680-4. *nature.com*. [En línea] 2006. doi:10.1038/sj.onc.1209954 <http://www.nature.com/onc/journal/v25/n51/full/1209954a.html>.
12. NCBI. BIRC3 WGS. [En línea] [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_018922.2?report=genbank&from=102071142&to=102093097](https://www.ncbi.nlm.nih.gov/nuccore/NC_018922.2?report=genbank&from=102071142&to=102093097).
13. —. Homo sapiens chromosome 11. [En línea] [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_000011.10?report=genbank&from=102317450&to=102339403](https://www.ncbi.nlm.nih.gov/nuccore/NC_000011.10?report=genbank&from=102317450&to=102339403).
14. —. POT1 protection of telomeres 1. [En línea] <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=25913>.
15. *POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia*. Al, Andrew J Ramsay et. s.l. : Nature, 2013, Nature Genetics. doi:10.1038/ng.2584.
16. GenBank. NCBI. [En línea] [https://www.ncbi.nlm.nih.gov/nuccore/NG\\_029232.1?report=genbank&from=5001&to=112598](https://www.ncbi.nlm.nih.gov/nuccore/NG_029232.1?report=genbank&from=5001&to=112598).
17. —. NCBI. [En línea] [https://www.ncbi.nlm.nih.gov/nuccore/NC\\_000007.14?report=genbank&from=124822386&to=124929983&strand=true](https://www.ncbi.nlm.nih.gov/nuccore/NC_000007.14?report=genbank&from=124822386&to=124929983&strand=true).
18. M, Agostino. *Introduction to the BLAST Suite and BLASTN. Practical Bioinformatics*. s.l. : Garland Science, 2012. pág. 396.

19. L., Bordoli. Similarity Searches on Sequence Databases: BLAST, FASTA. *Swiss Institute of Bioinformatics*. [En línea] 2003.
20. Handbook, The NCBI. National Center for Biotechnology Information (US). [En línea] [www.ncbi.nlm.nih.gov/books/NBK143764/..](http://www.ncbi.nlm.nih.gov/books/NBK143764/)
21. USCS. USCS Genome Browser. [En línea] <http://genome.ucsc.edu/FAQ/FAQblat.html>.
22. NIH. Genetics Home Reference. [En línea] <https://ghr.nlm.nih.gov/gene/XIAP>.
23. —. NIH U.S. National Library of Medicine. *Genetics Home Reference*. [En línea] <https://ghr.nlm.nih.gov/gene/GRIA3>.
24. «Natural occurrence». *Food Reviews International* 14 (2 & 3). K, Ninomiya. <http://www.tandfonline.com/doi/abs/10.1080/87559129809541157>, págs. 177-211.
25. Medicine, NIH U.S. National Library of. *Genetics Home Reference*. [En línea] <https://ghr.nlm.nih.gov/gene/DOCK7>.
26. Medicine, NIH U.S. National Library of. *Genetics Home Reference*. [En línea] <https://ghr.nlm.nih.gov/gene/SCP2>.
27. Al, Santos FR et. MedLine. [En línea] <https://www.ncbi.nlm.nih.gov/pubmed/10655552>.
28. Database, GeneCards Human Gene. *GeneCards*. [En línea] [http://www.genecards.org/cgi-bin/carddisp.pl?id\\_type=hgnc&id=9243](http://www.genecards.org/cgi-bin/carddisp.pl?id_type=hgnc&id=9243).
29. NCBI. Genome Reference Consortium. [En línea] <https://www.ncbi.nlm.nih.gov/grc/human>.
30. Europe, Protein Data Bank in. [En línea] <https://www.ebi.ac.uk/pdbe/entry/pdb/3t6p>.
31. Bank, PDB Protein Data. [En línea] <http://www.rcsb.org/pdb/explore/jmol.do?structureId=3KJP&bionumber=1>.
32. Guerra, J.I. Elejalde. *Anales de la Medicina Interna. Estrés oxidativo, enfermedades y tratamientos antioxidantes*. [En línea] <http://scielo.isciii.es/pdf/ami/v18n6/revision1.pdf>.
33. Gemma Viñasa, b, Teresa Puigb,c, Rut Porta. Estrés oxidativo en pacientes con cáncer: dos caras de una misma moneda. *ELSEVIER*. [En línea] <http://www.elsevier.es/es-revista-medicina-clinica-2-articulo-estres-oxidativo-pacientes-con-cancer-S0025775311011730>.
34. [En línea] <http://www.genecards.org/cgi-bin/carddisp.pl?gene=OXSR1&keywords=OXSR1>.
35. Roche, Instituto. Detección de mutaciones. [En línea] [https://www.instituto-roche.es/geneticaycancer/area\\_molecular/tecnicas\\_de\\_diagnostico\\_molecular/deteccion\\_de\\_mutaciones\\_moleculares.html](https://www.instituto-roche.es/geneticaycancer/area_molecular/tecnicas_de_diagnostico_molecular/deteccion_de_mutaciones_moleculares.html).
36. Cols., Robert Belshaw et. Long-term reinfection of the human genome by endogenous retroviruses. *PNAS*. [En línea] <http://www.pnas.org/content/101/14/4894>.
37. *Alu Repeats and Human Genomic Diversity. Nature Reviews: Genetics* 3. L., Batzer M. A. y Deininger P. 2002, págs. 370-379.
38. NIH. [En línea] <https://salud.nih.gov/articulo/vih-y-sida/>.
39. —. [En línea] <https://infosida.nih.gov/understanding-hiv-aids/glossary/3659/linfocito-t-cd4>.
40. ROBERT C. GALLO, et Cols. *Nature. Complete nucleotide sequence of the AIDS virus, HTLV-III*. [En línea] <https://www.nature.com/articles/313277a0>.

41. Walker, T. Stuart. *Microbiología*. s.l. : Mc Graw-Hill Internacional., 1998.
42. NIH, U.S. National Library of Medicine. Genetics Home Reference. [En línea]  
<https://ghr.nlm.nih.gov/gene/TP53>.

# ANEXO I

## Homología entre el gen OXSR1 y VIH-1

A continuación, se muestran el resultado de la búsqueda del gen OXSR1 y su secuencia:

### OXSR1 oxidative stress responsive 1 [ *Homo sapiens* (human) ]

Gene ID: 9943, updated on 7-May-2017

**Summary**

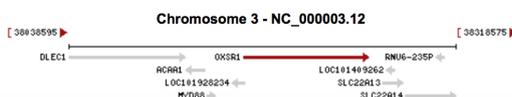
**Official Symbol** OXSR1 provided by HGNC  
**Official Full Name** oxidative stress responsive 1 provided by HGNC  
**Primary source** HGNC:HGNC:8508  
**See related** Ensembl:ENSG00000172939 MIM:604046; Vega:OTTHUMG00000131084  
**Gene type** protein coding  
**RefSeq status** REVIEWED  
**Organism** *Homo sapiens*  
**Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo  
**Also known as** OSR1  
**Summary** The product of this gene belongs to the Ser/Thr protein kinase family of proteins. It regulates downstream kinases in response to environmental stress, and may play a role in regulating the actin cytoskeleton. [provided by RefSeq, Jul 2008]  
**Orthologs** [mouse](#) [all](#)

**Genomic context**

**Location:** 3p22.2 [See OXSR1 in Genome Data Viewer Map Viewer](#)

**Exon count:** 22

Annotation release	Status	Assembly	Chr	Location
<a href="#">108</a>	current	GRCh38.p7 ( <a href="#">GCF_000001405.33</a> )	3	NC_000003.12 (38165051..38255488)
<a href="#">105</a>	previous assembly	GRCh37.p13 ( <a href="#">GCF_000001405.25</a> )	3	NC_000003.11 (38207022..38296979)



**Genomic regions, transcripts, and products**

**Genomic Sequence:** NC\_000003.12 Chromosome 3 Reference GRCh38.p7 Primary Assembly [Go to reference sequence details](#)

**Go to nucleotide:** [Graphics](#) [FASTA](#) [GenBank](#)

Genomic browser view showing the OXSR1 gene structure and associated transcripts. The OXSR1 gene is highlighted in green. Transcripts shown include XP\_011532633.1, XR\_001748396.1, XR\_017007601.1, XR\_001748397.1, XR\_017007602.1, XR\_005265638.2, NP\_005109.2, XR\_940733.1, XR\_245208.3, XR\_940731.1, XR\_940728.1, XR\_001740658.1, XR\_940732.2, XR\_940729.2, XR\_940730.2, NP\_004256.3, XR\_001748370.1, and NP\_017007519.1.

Haciendo la consulta en el enlace GenBank, se puede observar la secuencia de nucleótidos asociada al gen OXSR1 es:

GenBank - Send: ▾

### Homo sapiens chromosome 3, GRCh38.p7 Primary Assembly

NCBI Reference Sequence: NC\_000003.12

[FASTA](#) [Graphics](#)

LOCUS NC\_000003 90438 bp DNA linear CON 06-JUN-2016

DEFINITION Homo sapiens chromosome 3, GRCh38.p7 Primary Assembly.

ACCESSION NC\_000003 REGION: 38165051..38255488 GPC\_000001295

VERSION NC\_000003.12

DBLINK BioProject: [PRJNA168](#)  
Assembly: [GCF\\_000001405.33](#)

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
Catarrhini; Hominidae; Homo.

**Change region shown**

Whole sequence (abbreviated view)

Selected region

from: 38165051 to: 38255488 Update View

**Customize view**

Abbreviated view

Customize

**Basic Features**

Default features

Gene, RNA, and CDS features only

**Display options**

Show sequence

Como se observa en la imagen, la secuencia codificante de este gen comprende la región 38.165.051 bp hasta 38.255.488 bp, con una longitud total de 90.438 bp en el cromosoma 3.

A continuación, se realiza la búsqueda de homólogos, y se muestra el resultado de BLASTN. (Se puede consultar la secuencia completa aquí: <https://goo.gl/9XEWIK>):

**Query ID** NC\_000003.12 **Database Name** nr

**Description** Homo sapiens chromosome 3, GRCh38.p7 Primary Assembly **Description** Nucleotide collection (nr)

**Molecule type** dna **Program** BLASTN 2.6.1+ [Citation](#)

**Query Length** 90437

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

**Graphic Summary**

Distribution of the top 200 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------

**Descriptions**

**Sequences producing significant alignments:**

Select: [All](#) [None](#) [Selected:0](#)

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Homo sapiens clone AD-13 HIV-1 integration site	372	23292	12%	4e-99	90%	<a href="#">AY255031.1</a>
<input type="checkbox"/> HIV-1 isolate M0107v1d14-5M15-C1 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982185.1</a>
<input type="checkbox"/> HIV-1 isolate M0107v1d14-5M12-H12 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982179.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M12-C07 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982067.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M05-B08 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982053.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M05-B05 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982052.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M05-B04 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982051.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M05-B03 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982050.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M03-E07 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982042.1</a>
<input type="checkbox"/> HIV-1 isolate C9003v1d14-5M01-A03 from USA genomic sequence	347	8601	11%	1e-91	87%	<a href="#">KT982027.1</a>
<input type="checkbox"/> Homo sapiens DNA, HIV-1 integration site, isolate:050204Pl.3#1	345	7854	10%	5e-91	87%	<a href="#">AB256413.1</a>
<input type="checkbox"/> HIV-1 isolate M0107v1d14-5M12-H07 from USA genomic sequence	343	8563	11%	2e-90	88%	<a href="#">KT982178.1</a>
<input type="checkbox"/> HIV-1 isolate M0102v1d14-5M27-G01 from USA genomic sequence	343	8208	11%	2e-90	86%	<a href="#">KT982173.1</a>
<input type="checkbox"/> Homo sapiens DNA, HIV-1 integration site, isolate:011102PL.2#30	333	9086	11%	3e-87	86%	<a href="#">AB256280.1</a>
<input type="checkbox"/> Homo sapiens DNA, HIV-1 integration site, isolate:050204Pl.3#276	315	6913	11%	9e-82	85%	<a href="#">AB256463.1</a>
<input type="checkbox"/> Homo sapiens DNA, HIV-1 integration site, isolate:011102PL.2#238	313	8059	11%	3e-81	84%	<a href="#">AB256303.1</a>
<input type="checkbox"/> Homo sapiens clone AD-11 HIV-1 integration site	313	13992	12%	3e-81	83%	<a href="#">AY255029.1</a>
<input type="checkbox"/> Homo sapiens clone AD-5 HIV-1 integration site	307	12004	11%	1e-79	85%	<a href="#">AY255023.1</a>
<input type="checkbox"/> Homo sapiens clone AL-7 HIV-1 integration site	298	7092	10%	7e-77	82%	<a href="#">AY253274.1</a>

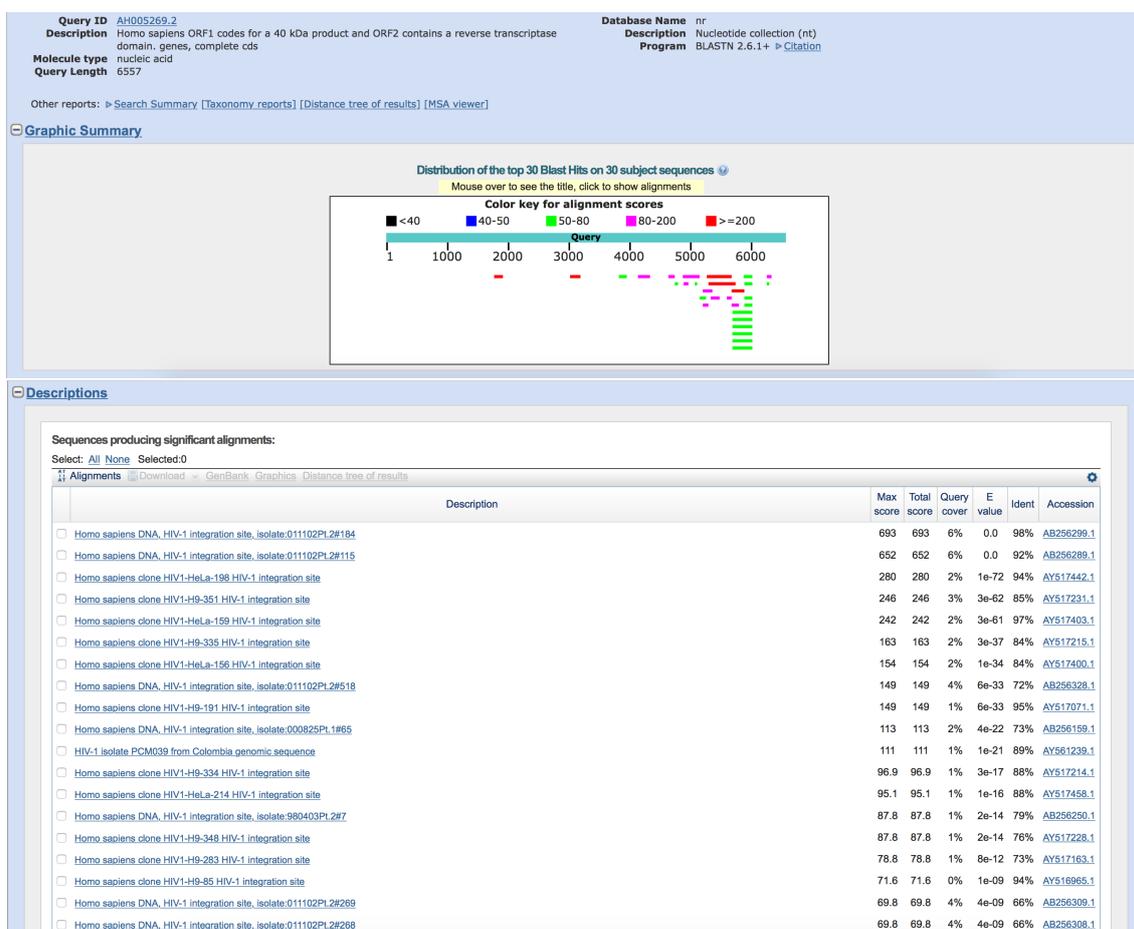
Como se observa, BLASTN nos devuelve al menos más de 100 resultados, con unos valores de indentidad que rondan entre el 80 y 90% y con un *E-Value* máximo de  $4e-99$ .

## ANEXO II

### Homología entre el gen ORF2 y VIH-1

Para el estudio de la homología entre el gen ORF2 y VIH-1, se ha tomado como secuencia la anotación GenBank con número de acceso: AH005269.2.

Utilizando la herramienta BLASTN disponible en el portal NCBI, y usando las secuencias disponibles de VIH-1 en la base de datos nr/nt, se obtienen los siguientes resultados:



El resultado devuelve homologías en regiones del HIV-1 identificadas como "Integration site", al igual que los resultados anteriores. Se obtienen valores de *Identity* de 98% y un E-Value de 0.0.

Los resultados de esta búsqueda, son menores que los anteriores en comparación con los otros genes de estudio. La secuencia completa se puede consultar aquí: <https://goo.gl/AN1O3K>