Variantes de *splicing*: detección y su efecto en el mRNA

Claudia Vázquez Calvo

Máster universitario en Bioinformática y bioestadística UOC-UB
Área 4, Subárea 1: Estudios genéticos de enfermedades humanas

Helena Brunel Montaner
Antoni Pérez Navarro

05 de enero de 2021

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Variantes de splicing: detección y su efecto en el mRNA*
Nombre del autor:	*Claudia Vázquez Calvo*
Nombre del consultor/a:	*Helena Brunel Montaner*
Nombre del PRA:	*Antoni Pérez Navarro*
Fecha de entrega (mm/aaaa):	01/2021
Titulación::	*Máster universitario en Bioinformática y bioestadística UOC-UB*
Área del Trabajo Final:	*Área 4, Subárea 1: Estudios genéticos de enfermedades humanas*
Idioma del trabajo:	*Castellano*
Palabras clave	*Splicing, mutaciones, NGS*

Resumen del Trabajo (máximo 250 palabras):

El *splicing* es un proceso que ocurre entre la transcripción y la traducción y consiste en eliminar las regiones no codificantes del mRNA para producir un transcrito maduro. Este proceso es necesario para que la traducción se realice correctamente. La presencia de mutaciones puntuales que alteran las secuencias reguladoras conduce a *splicing* aberrante, pudiendo producir trastornos monogénicos hereditarios específicos.

Las mutaciones que afectan al *splicing* se han clasificado en cinco tipos, según la posición que ocupan en el DNA y según el efecto que producen en él.

En el presente TFM, se pretende estudiar algunos de los diferentes tipos de predictores de *splicing in silico*, partiendo inicialmente de 30; así como comprobar la exactitud con la que éstos predicen el efecto que pueden producir dichas mutaciones y clasificarlas según los cinco tipos fijados.

Se ha estudiado un conjunto de 99 variantes, entre las cuales de unas se conocía el efecto que producen en el *splicing*, y otras se obtuvieron de una base de datos de enfermedades humanas, no conociendo su efecto. Se ha hecho un análisis estadístico y descriptivo de cada uno de los predictores para determinar el de mayor exactitud y los problemas que tienen cada uno de ellos.

Sólo unos pocos predictores dan resultados que pueden ser clasificables. Otros no pueden ser extrapolables a los tipos de *splicing* estudiados o generan resultados que pueden llevar a concluir tipos distintos. Al ser predicciones, es imprescindible validar experimentalmente para poder determinar el verdadero efecto de estas mutaciones en el *splicing.*

Abstract (in English, 250 words or less):

Splicing is a process that occurs between transcription and translation and consists in eliminating the non-coding regions of the mRNA to produce a mature transcript. This process is necessary for the translation to be successful. Presence of point mutations that alter regulatory sequences leads to aberrant splicing, which can produce specific hereditary monogenic disorders.

Mmutations that affect splicing have been classified into five types, according to the position they occupy in the DNA and according to the effect they produce on it.

In this *TFM*, it is intended to study some of the different types of splicing predictors *in silico*, starting initially from 30; as well as to verify the accuracy they have to predict the effect that these mutations can produce and to classify them intro the five fixed effects.

A set of 99 variants has been studied, for some the effect they produce on splicing was known, and others were obtained from a human disease database, without knowing their effect. A statistical and descriptive analysis has been performed for each of the predictors to determine the one with the highest accuracy and the problems that each one of them have.

Only a few predictors give results that can be classifiable. Others cannot be extrapolated to the types of splicing studied or generate results that can lead to the conclusion of different types. As they are predictions, it is essential to validate experimentally in order to determine the true effect of these mutations on splicing.

Índice

Lista de figuras

Lista de tablas

1. Introducción

1.1 Contexto y justificación del Trabajo

1.1.1 *Splicing*: fundamentos

El dogma central de la biología molecular establece que el DNA contiene las instrucciones necesarias para producir una proteína, que son copiadas a un RNA mensajero, en un proceso llamado transcripción, y, más tarde, el RNA usa esas instrucciones para producir una proteína, proceso también llamado traducción (Crick, 1970). En células eucariotas, el mRNA tiene que transportarse desde el núcleo de la célula hasta el citoplasma para llevar a cabo la traducción. Antes de que se produzca este transporte, el mRNA tiene que llevar a cabo un proceso de maduración en el que se retiran las regiones no codificantes, o intrones, para que el mensajero solo contenga las regiones codificantes, o exones, que son aquellas que se van a traducir a proteína. Este proceso de maduración se conoce como *splicing* (Rahman *et al*., 2015).

El *splicing* es un proceso altamente controlado que es llevado a cabo por una maquinaria de *splicing* del núcleo, llamada *spliceosome*. El *spliceosome* humano central, junto con los factores reguladores asociados, comprende más de 300 proteínas y cinco RNA nucleares pequeños (snRNA). La arquitectura del *spliceosome* se somete a una remodelación dinámica antes, durante y después de la reacción de *splicing*. Este elimina los intrones y une los exones para generar una molécula de mRNA madura. La maquinaria se ensambla en la molécula de pre-mRNA en secuencias específicas ubicadas en los límites exón-intrón y que definen los sitios de *splicing* (SS) 3' y 5' y el *branch point site* (BPS). Además del *spliceosome*, hay una serie de proteínas reguladoras *trans* que participan en la modulación de la reacción de *splicing* y que actúan como activadoras o represoras del proceso al unirse a elementos potenciadores o silenciadores exónicos o intrónicos (Urbanski *et al.*, 2018). A estos elementos se les conoce como elementos reguladores *cis*, entre los que se incluyen los *exonic* e *intronic splicing enhancers* (o por sus siglas en inglés ESE e ISE, respectivamente) y los *exonic* e *intronic splicing silencers* (o por sus siglas en inglés ESS e ISS, respectivamente) (Glisovic *et al.*, 2008).

Además, se debe tener en cuenta la existencia del fenómeno de *splicing* alternativo, el cual es un paso clave de la regulación de la expresión génica postranscripcional. Contribuye a la diversidad proteómica y funcional al permitir la producción de distintas isoformas de RNA a partir de un solo gen. El *splicing* alternativo proporciona plasticidad transcripcional al controlar qué isoformas de RNA se expresan en un momento dado en un tipo celular concreto. Las células cancerígenas subvierten este proceso para producir isoformas que benefician la proliferación o migración celular, o para que no puedan escapar de la muerte celular (Biamont *et al.,* 2014). Se estima que alrededor del 60% de los genes humanos sufren *splicing* alternativo (Modrek & Lee, 2002). Un tercio de los eventos de *splicing* alternativos introducen codones de terminación prematura

(*premature termination codons*, o por sus siglas en inglés PTC), que provocan la degradación del mRNA por desintegración mediada sin sentido (*nonsense-mediated decay*, o por sus siglas en inglés NMD). Por lo tanto, la regulación del *splicing* alternativo controla la expresión temporal y espacial de isoformas funcionalmente diversas, la regulación intermitente por NMD u otras respuestas reguladoras postranscripcionales (Lewis *et al.*, 2003).

Cualquier error durante el proceso de *splicing* puede conducir a una eliminación inadecuada del intrón y, por lo tanto, provocar alteraciones del marco de lectura abierto (*open reading frame*, o por sus siglas en inglés ORF). El complejo de *spliceosome* debe reconocer y cortar correctamente las secuencias intrónicas de la molécula de pre-mRNA. La identificación adecuada del sitio de *splicing* es complicada, ya que las secuencias consenso son muy cortas y hay muchas otras secuencias similares a los motivos consenso de los sitios de *splicing* canónicos. Estas secuencias se conocen como secuencias crípticas, no canónicas o de sitios de *pseudo-splicing* (Cartegni *et al.*, 2002). El complejo deberá diferenciar entre ellas para no provocar fallos funcionales en la célula al generar un mRNA incorrecto.

1.1.2 Mutaciones en el DNA: tipos y estudio

Una mutación es un cambio en una secuencia genética. Las mutaciones incluyen cambios tan pequeños como la sustitución de un solo bloque de construcción de DNA, o base nucleotídica, por otra base nucleotídica. Algunas mutaciones son hereditarias porque se transmiten a la descendencia de un parental que lleva una mutación a través de la línea germinal, es decir, a través de un óvulo o un espermatozoide portador de la mutación. También existen mutaciones no hereditarias que ocurren en células fuera de la línea germinal, que se denominan mutaciones somáticas. Algunas mutaciones no producen cambios en la secuencia de aminoácidos de la proteína codificada y pueden describirse como mutaciones silenciosas. Otras mutaciones dan como resultado productos proteicos anormales (Nature, 2018).

Cuando se estudia la secuenciación de un paciente, se pueden identificar cambios en éste con respecto al genoma de referencia. Lo primero que se debe hacer es determinar si se trata de un cambio novel (***Fig. 1***). Para ello, se lleva a cabo una búsqueda por las bases de datos genómicos y de mutaciones patológicas, como pueden ser *ENSEMBL* (Zerbino, *et al.,* 2018), *HGMD* (Stenson, *et al.,* 2020), o *gnomAD* (Koch, 2020), entre otras. Necesariamente si un cambio tiene una frecuencia alta en población control, difícilmente va a poder ser una mutación patológica. Éstas habitualmente son noveles o tienen frecuencias muy bajas. De hecho, la definición de polimorfismo se basa en la presencia de un cambio en población sana con una frecuencia superior al 1%. Adicionalmente, estas bases de datos darán información previa o bibliografía a consultar, útil para proseguir con el estudio del cambio.

Es importante también la realización de estudios genéticos (***Fig. 1***). Por un lado, se debe comprobar que el cambio candidato a ser mutación clínica lo

tiene el individuo objeto de estudio, ya que podría tratarse de un falso positivo. Esto se realizaría amplificando la región de interés con una PCR y posterior secuenciación por Sanger. Para saber si este cambio cosegrega con la enfermedad, se debe comprobar si familiares del probando tienen el cambio si están enfermos, o no lo tienen estando sanos. Esto es lo que se conoce como análisis de segregación y es la forma más sencilla para descartar cambios candidatos, siempre y cuando los individuos estudiados hayan sido evaluados clínicamente por el facultativo clínico (Vázquez, 2019).

Un tercer paso a realizar comprende los estudios *in silico* (***Fig. 1***). En la web existen múltiples softwares que predicen si un cambio de aminoácido es patológico como *SIFT* (Ng & Henikoff, 2003) o *PolyPhen-2* (Adzhubei *et al.*, 2013), o si un cambio puede conducir a un *splicing* anómalo como *NNSPLICE versión 0.9* (Reese *et al.*, 1997) o *NetGene2* (Hebsgaard *et al.*, 1996; Brunak *et al.*, 1991). También el estudio de la conservación del residuo es útil, ya que un aminoácido no conservado es probablemente poco relevante. Para ello se han diseñado herramientas tales como *PhyloP* (Pollard *et al.*, 2010) o *GERP* (Cooper *et al.*, 2005). Entre las aproximaciones computacionales disponibles está el análisis de las consecuencias en la estructura de la proteína, empleando programas tales como *Coot* (Emsley & Cowtan, 2004), si bien para este tipo de estudios es imprescindible contar con el cristal de la proteína en la *Protein Data Base* (PDB) (Berman, *et al.,* 2000). En su conjunto, todas estas aproximaciones son muy económicas y sencillas de aplicar, aunque no dejan de ser predicciones o estimaciones de lo que pudiera estar ocurriendo.

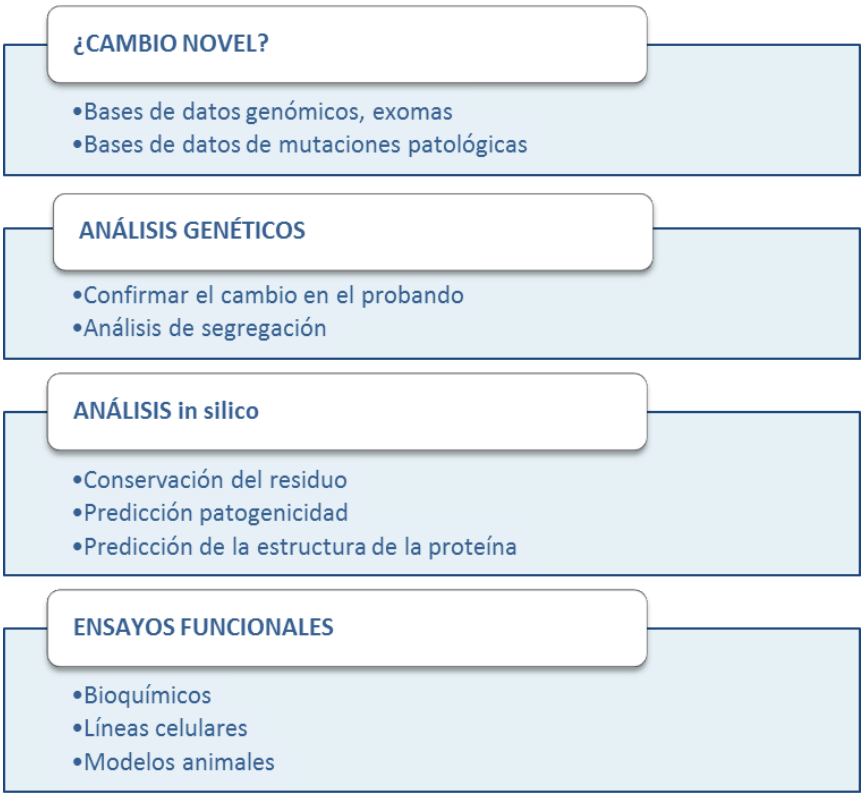

Figura 1. *Esquema de los pasos a seguir para establecer que una variante es patológica (Vázquez, 2019).*

Por último, también se pueden realizar ensayos funcionales para ganar evidencias experimentales sobre la patogenicidad del cambio. Éstos son especialmente relevantes cuando se trata de cambios noveles, máxime si el gen nunca ha sido previamente asociado a enfermedad humana. Analizando 84 variantes, descritas como responsables de fibrosis quística, con análisis de expresión se demostró que 6 de ellas (13%) realmente no afectaban a la actividad de la proteína mutada, habiendo sido predichas como deletéreas o probablemente deletéreas (Raraigh *et al.*, 2018). Este hallazgo subraya la necesidad de ser cautos con las herramientas *in silico* empleadas habitualmente en el análisis de las consecuencias de las mutaciones.

1.1.3 Variantes de *splicing*: tipos y efectos

El *splicing* de RNA desregulado o anormal se ha encontrado muy comúnmente en las células cancerosas y está relacionado con la carcinogénesis (David & Manley, 2010), aunque análisis revelaron que la diferencia entre genes con *splicing* alternativo es en realidad levemente menor en tumores comparados con tejidos normales (Kim *et al.*, 2008). Hay muchos estudios que se han dedicado a analizar los diferentes cambios que pueden ocurrir en el DNA y qué efecto tienen sobre el *splicing*, así como analizar muchas otras mutaciones sin influencia en él, la mayoría de las cuales se encuentran anotadas en bases de datos como *COSMIC* o *TCGA* (Tate *et al.*, 2019; Wang *et al.*, 2016). Estudiar la importancia de las mutaciones de *splicing* en la patogénesis de las enfermedades genéticas puede dar lugar a numerosos estudios experimentales y clínicos que se centran en el desarrollo de fármacos, los cuales puedan revertir el efecto de estas mutaciones (Abramovicz & Gos, 2018).

Alteraciones en el *splicing* del RNA están implicadas en varios tipos de enfermedades humanas a través de la interrupción de elementos *cis* dentro de los genes afectados o de factores *trans*, que son necesarios para que se produzca el *splicing* normal o la regulación del proceso. En general, la alteración de los elementos *cis* afecta solo a un gen, mientras que defectos en los factores *trans* podrían afectar a múltiples genes (Wang & Cooper, 2007).

Las alteraciones de los elementos *cis* son un mecanismo bien conocido que causa la traducción anormal de proteínas. Se estima que hasta el 50% de las mutaciones relacionadas con el *splicing* que causan enfermedades son mutaciones que actúan en elementos *cis*, lo que conduce a la traducción de proteínas alteradas (Cartegni *et al.*, 2002). Por lo general, este tipo de mutaciones dan como resultado o el *skipping* (pérdida o salto) del exón completo o de un fragmento de dicho exón. Si la deleción resultante se da dentro del marco de lectura abierto, se sintetizará una proteína más corta (es decir, la misma proteína que se habría sintetizado sin la parte que se pierde). Sin embargo, cuando la deleción da como resultado un cambio del ORF, se puede introducir un codón de parada prematuro (PTC) y, en caso de que se lleve a cabo la traducción, se produciría una proteína más corta (ya que la introducción del PTC impediría que se tradujera el resto de la proteína a partir

de él, reduciendo el tamaño de ésta). No obstante, la presencia de un PTC en la transcripción generalmente conduce a una degradación más rápida del mRNA durante el proceso protector de *non-sense mediated decay* (NMD). La degradación del RNA mensajero defectuoso evita la síntesis de proteínas aberrantes y tiene el mismo efecto que la deleción de genes o la mutación sin sentido (Sterne-Weiler & Sanford, 2014).

Existen otros cambios que pueden afectar a elementos *cis*, como los cambios que afectan al *brach point site* (BPS) o a las secuencias de *polypyrimidine tract*, que se unen a proteínas específicas involucradas en la formación del complejo de *splicing*. Mutaciones localizadas en el BPS podrían dar lugar a un exón *skipping*, debido a la unión inadecuada de las proteínas de *splicing* SF1 y U2 snRNP y la interrupción del sitio del *acceptor* natural. También pueden causar retención de intrones (en su totalidad o un fragmento) si crean un nuevo sitio de *splicing* 3' (Caminsky *et al*., 2014). Por otro lado, cualquier mutación en las secuencias de *polypyrimidine tract* probablemente lleve a alteraciones de *splicing*, aunque la lista de tales variantes es limitada (Ward & Cooper, 2010).

A parte de las alteraciones en elementos *cis*, también puede haber mutaciones que afecten al *splicing* en los elementos *trans*, en las que la mutación de la línea germinal afecta a los principales componentes del *spliceosome*, se han reconocido como las causas de las formas autosómicas dominantes de retinosis pigmentaria (Cooper *et al.*, 2009). Este tipo de mutaciones son más minoritarias y menos conocidas.

1.1.4. Algoritmos de detección de las variantes de *splicing*

Como se ha comentado en el apartado 1.1.2, las variantes de *splicing* son mutaciones del DNA y tienen que pasar por los puntos mencionados en ese apartado para poder determinar el efecto que realizan en el fenotipo del gen en el que se encuentran. Este TFM se centra en una aproximación bioinformática al problema de la detección de variantes de *splicing*. Durante las últimas décadas, se han generado predictores *in silico* o algoritmos bioinformáticos que pretenden aproximar el comportamiento de las mutaciones de *splicing*. Los algoritmos existentes para el análisis de *splicing* difieren entre sí en la base de datos, que contiene información sobre las secuencias consenso; el modelo estadístico utilizado para el análisis o los métodos de entrenamiento que se emplean en los enfoques de *machine learning*. La mayoría de las herramientas se centran en el análisis de los sitios consenso de *splicing* y requieren como *input* la secuencia a estudiar (Abramovicz & Gos, 2018). Existen diferentes tipos de predictores que se encargan de analizar distintos aspectos del *splicing*. Hay herramientas que se basan en el modelo de position matriz de pesos, el modelo probabilístico de *maximum dependence decomposition*, técnicas de machine learning o el modelo de distribución de máxima entropía. Hay otros que estudian como una mutación distante puede afectar al *splicing* o para predecir el *exon skipping*, la activación del sitio críptico o la generación de transcritos aberrantes a partir de la secuencia primaria. También pueden estudiar como un SNP puede afectar al *brach point site* o a las secuencias de *polypyrimidine tract* o predecir la patogenicidad de la variante, así como

estudiar posibles alteraciones en los ESE o ESS o analizar el potencial de las secuencias ESE/ESS (Abramovicz & Gos, 2018).

Sin embargo, se debe tener en cuenta que estas herramientas son predictores, es decir, no indican exactamente el efecto que producen en la realidad estas variantes. Para saber con certeza qué consecuencia van a tener en el *splicing*, solo se puede comprobar llevando a cabo ensayos funcionales. Por ejemplo, se puede realizar un RT-PCR en el que se amplifique la región donde se supone que puede estar afectando ese cambio al *splicing* y observar con una simple electroforesis qué consecuencia tiene en el mRNA. El principal problema con este enfoque es la posibilidad de que se active el mecanismo *non-mediated decay* (NMD). En tal situación, el efecto de la posible mutación de *splicing* se puede perder fácilmente. Para superar esta desventaja, las células del paciente pueden tratarse con inhibidores de la NMD como la puromicina, que bloquea la degradación del RNA (Baralle & Baralle, 2005).

Otra posibilidad es hacer ensayos con minigenes. Este sistema es especialmente útil para el análisis de genes con bajo nivel de expresión en leucocitos o fibroblastos (Singh & Cooper, 2006). En este ensayo, el fragmento amplificado del gen analizado se clona en un plásmido de expresión especial, que permite el análisis del *splicing* de pre-mRNA. Este enfoque se puede utilizar para confirmar que la variante de *splicing* potencial afecta la eficiencia del *splicing* o causa la activación de los sitios de *splicing* críptico alternativos, y para probar el papel de elementos que actúan en *cis* en la regulación del *splicing* (Sharma *et al.*, 2014).

Por lo tanto, los predictores agilizan gran parte del trabajo de análisis, pero no son, todavía, sustituyentes de los ensayos experimentales, sino complementarios a éstos.

1.2 Objetivos del Trabajo

El objetivo principal de este trabajo es estudiar los diferentes algoritmos bioinformáticos o predictores *in silico* para variantes genéticas que afectan al *splicing* dentro un conjunto de variantes extraídas de un experimento de NGS y estudiar su nivel de precisión.

Para llevar a cabo este objetivo será necesario realizar los siguientes objetivos secundarios:

- Hacer una revisión bibliográfica de métodos *in silico* de predicción de *splicing.*
- Descargar una base de datos anotada de variantes genéticas, al ser posible relacionada con enfermedades humanas.
- Interpretar la información de la anotación de variantes y extraer la información necesaria para averiguar si una variante afecta al *splicing*.
- Definir formalmente los criterios mediante los cuales una variante genética puede afectar al *splicing*.

- Comparación entre predictores *in silico* de las variantes seleccionadas al azar.
- Realizar un análisis descriptivo y estadístico de las predicciones obtenidas.

1.3 Enfoque y método seguido

Existen diferentes tipos de variantes que pueden afectar el *splicing*, por lo que el enfoque del trabajo consistirá en caracterizar cada uno de estos tipos de variantes según las anotaciones. El enfoque de trabajo consiste en, desde una base de datos, reducirla hasta quedarnos solo con las mutaciones que producen un cambio de nucleótidos puntual (43 de los 47 millones iniciales).

A continuación, se ha aplicado un algoritmo que consigue separar en tres ficheros distintos estas variables según los tipos previamente fijados: tipo I o IV, tipo II y tipo III o V. Con los datos proporcionados de la base de datos y sólo con la anotación a nivel de DNA y sin la secuencia adyacente, no se puede distinguir entre tipo I o V y entre tipo III o V, dado que esto requiere analizar la secuencia adyacente y si este cambio produce un efecto significativo en la aparición o no de regiones necesarias para que se produzca correctamente el *splicing*.

Una vez separados los ficheros, se seleccionará una serie de variables al azar de los tres y se analizarán por diferentes predictores *in sillico* que se encuentran en la web para comparar los resultados entre sí. Se ha optado por predictores que tienen portal web en vez de otro tipo de programas descargables (como de GitHub) por su comodidad a la hora de introducir las variables.

Una vez se tenga las predicciones de estas variantes, se analizará la efectividad de cada uno de los predictores, obteniendo los falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos de cada predictor.

1.4 Planificación del Trabajo

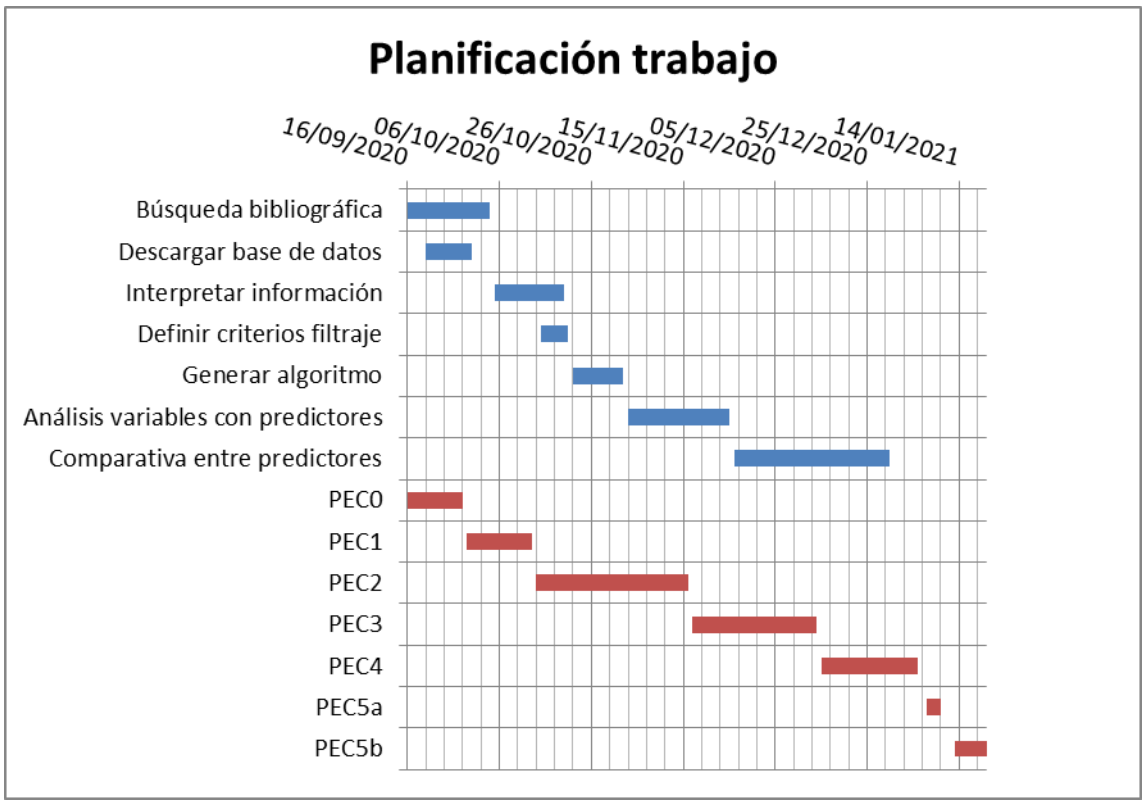

Figura 2. *Diagrama de Gantt que explica las tareas a realizar para el presente TFM, así cómo el tiempo dedicado a cada una de ellas, comparando con las diferentes entregas a presentar.*

1.5 Breve sumario de productos obtenidos

Con este trabajo, se van a obtener una serie de productos. Estos productos serán:

- La caracterización de los criterios de clasificación de las variantes y su definición formal, es decir, transformar las características de cada uno de los tipos de variables (Wimmer *et al.,* 2007) en términos de las anotaciones de la base de datos.
- Las comparativas entre los diferentes predictores *in silico.*
- Los resultados de los análisis de las variantes obtenidas que puedan aportar información relevante al ámbito de la anotación de variantes de *splicing*.
- Los resultados del análisis estadístico para cada predictor.
- Clasificación según los tipos de variables (Wimmer *et al.,* 2007) de cada predictor para cada variante.
- La memoria y la presentación final del trabajo realizado.

1.6 Breve descripción de los otros capítulos de la memoria

En la introducción del trabajo se pretenderá explicar con claridad el concepto de *splicing*, así como los antecedentes en el estudio de las variantes y su efecto en este fenómeno. Una vez expuestos lo anteriormente estudiado en el campo, en la parte central de la memoria se establecerán los criterios de filtrado de las variantes que permita diferenciarlas entre sí. Se expondrá cómo se genera el *script* en el lenguaje de programación de interés y los resultados obtenidos. A continuación, se mostrará cómo se ha seleccionado cada uno de los predictores para cada tipo de variantes y los resultados ue se obtienen para cada una de ellas. En la discusión, se analizarán los resultados obtenidos y se expondrá la efectividad de los predictores, así como cuál es el mejor para cada tipo de mutaciones.

2. Resto de capítulos

2.1 Materiales y métodos

2.1.1 Selección y estudio de base de datos

Para llevar a cabo este TFM, se necesitaron datos de mutaciones en el DNA. Existen diferentes tipos de bases de datos de proyectos de *Next Generation Sequencing* (NGS) que contienen mutaciones como *The Cancer Genome Atlas* (TCGA) (Wang *et al.*, & Zenklusen, 2016), *The Human Gene Mutation Database* (*HGMD*) (Stenson *et al.,* 2020) o gnomAD (Koch, 2020). Sin embargo, hay algunas bases de datos que no ofrecen toda la información de manera pública, como HGMD, en la es necesario tener una cuenta de pago para acceder a los datos. Por lo tanto, a la hora de elegir la base de datos que se empleó, se tuvo en cuenta que los datos totalmente fueran descargables fácil, pública y gratuitamente. En cuanto al formato del fichero, existen diferentes tipos donde se pueden guardar la información de las mutaciones. Uno de los más empleados es el formato *variant call format* (VCF), que se ha utilizado para diferentes tipos de proyectos de NGS como *1,000 Genomes* (1000 Genomes Project Consortium *et al.*, 2015), *the exome aggregation consortium* (ExAC) (Lek *et al.*, 2016), y *the cancer genome atlas* (TCGA) (Wang *et al.*, 2016). También se puede encontrar la información en formato *tab separated values* (TSV), como ocurre con la mayoría de los ficheros que se encuentran en la base de datos *Catalogue Of Somatic Mutations In Cancer* (COSMIC) (Forbes *et al.*, 2017). Estos tipos de ficheros cubren, en muchos casos, la totalidad del genoma humano, lo que hace que la información que contienen es del orden de millones de mutaciones. Descargarlos y manejarlos requiere herramientas específicas, como *Python*, *Bash* o *RStudio*.

En particular, se eligió la base de datos *Catalogue Of Somatic Mutations In Cancer* (COSMIC) (Forbes *et al.*, 2017) para poder seleccionar y analizar las variables. En esta base de datos, con una cuenta gratuita si se estudiante, se puede acceder a toda la información de la base de datos. Se ha reportado la presencia de alto porcentaje de mutaciones que afectan al *splicing* en genes relacionados con el cáncer. Por ejemplo, se ha observado que casi todas las mutaciones testadas en los exones 8 y 15 de gen MLH1 (relacionado con el Síndrome de Lynch) afectaban significativamente al *splicing* (100% y 71%, respectivamente) (Rhine *et al.,* 2018). Las células cancerosas expresan una tasa elevada de transcritos que contienen codones de *stop* prematuros (PTC), consistente con una tasa aumentada de *splicing* incorrecto en relación con los tejidos normales (Chen *et al.*, 2011). Por lo tanto, se estaría partiendo de un fichero con un elevado porcentaje de mutaciones afectando al *splicing*.

De entre todos los ficheros disponibles, se seleccionó aquel que contaba con las mutaciones anotadas en la nomenclatura recomendada por la *Human Genome Variation Society* (HGSV) (den Dunnen *et al.*, 2016). Esta nomenclatura, actualmente se reconoce como la nomenclatura estándar en el

diagnóstico molecular (Gulley *et al.*, 2007; Richards *et al.*, 2015). Este fichero, que está en extensión *.tsv*, se analizó mediante una línea de comandos. De este fichero, la información que se extrajo fue: el nombre del gen donde se encuentra la mutación, el transcrito en el que se encuentra la mutación, la posición cromosómica en la que se haya la mutación (sabiendo en qué versión del genoma está anotada la mutación) y la anotación HGSV (den Dunnen *et al.*, 2016) de la mutación.

2.1.2 Detección de variantes genéticas que afecta al *splicing*

Originalmente, este TFM pretendía proponer un único algoritmo capaz de llevar a cabo la detección de variantes que afectan al *splicing*, pero no ha sido posible llevarlo a cabo con los recursos computacionales disponibles y debido a la limitación de uso a 100 variantes de la principal herramienta. Por lo tanto, el objetivo final de este TFM consiste en comparar diferentes predictores de *splicing in silico* en un conjunto reducido de variantes seleccionado aleatoriamente a partir del fichero original.

La **figura 3** refleja, en un diagrama de flujo, los pasos realizados para la detección de variantes genéticas que afectan al *splicing*, que pasarán a ser explicados a continuación.

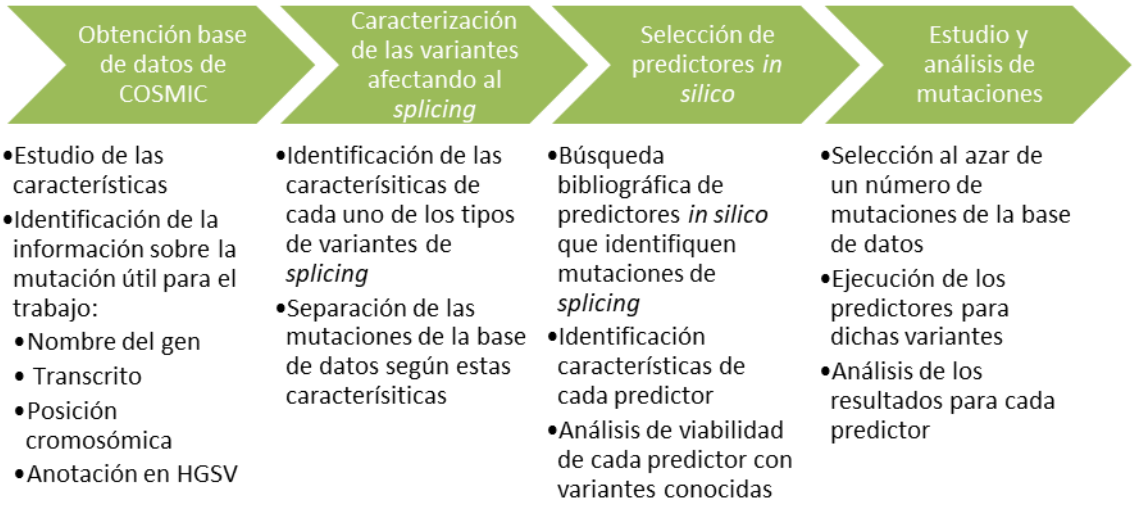

Figura 3. *Diagrama de flujo del procedimiento seguido para el análisis de mutaciones.*

El primer paso consistió en descargar la base de COSMIC. Una vez descargada, se localizó dónde se encontraba la información de interés: el nombre del gen donde se encuentra la mutación (columna 1), el transcrito en el que se encuentra la mutación (columna 2), la anotación (columna 20) y la posición cromosómica en la que se haya la mutación (sabiendo en qué versión del genoma está anotada la mutación) (columna 26).

El segundo paso consistió en detectar qué variantes de estas afectan al *splicing*. Para ello, primero se tuvo que encontrar las características que deben tener las variantes para poder estar alterando el correcto funcionamiento del *splicing*. Seguidamente, se separaron las mutaciones que consistían en cambios puntuales en archivos diferentes según los tipos identificados.

En el tercer paso del proceso, se aplicaron diferentes predictores *in silico* para la identificación del efecto de una mutación en el *splicing* y se observó cuáles son mejores para cada tipo de variante.

Seguidamente, se seleccionaron ejemplos al azar de mutaciones de la base de datos para ser analizados en diferentes predictores *in silico* para estudiar el posible efecto o no de estas variantes en este fenómeno.

Finalmente, como último paso, se analizaron los resultados obtenidos para estas mutaciones.

2.1.2.1. Caracterización de variantes

Según Wimmer *et al*. (2007), las mutaciones que pueden afectar al *splicing* se pueden clasificar en cinco tipos distintos (**Fig. 4**):

(I) Mutaciones dentro de los sitios de *splicing* canónicos que conducen a la omisión del exón completo.
(II) Variantes intrónicas profundas que crean nuevos sitios de *splicing* y dan como resultado la inclusión de exones crípticos.
(III) Variantes exónicas que dan como resultado la pérdida de un fragmento de exón.
(IV) Variantes en los sitios de *splicing* canónicos que dan como resultado el uso del sitio de *splicing* críptico exónico o intrónico y conducen a la inclusión de un fragmento de intrón o un salto de fragmento de exón, respectivamente.
(V) Mutaciones dentro del exón que conducen a la omisión del exón completo.

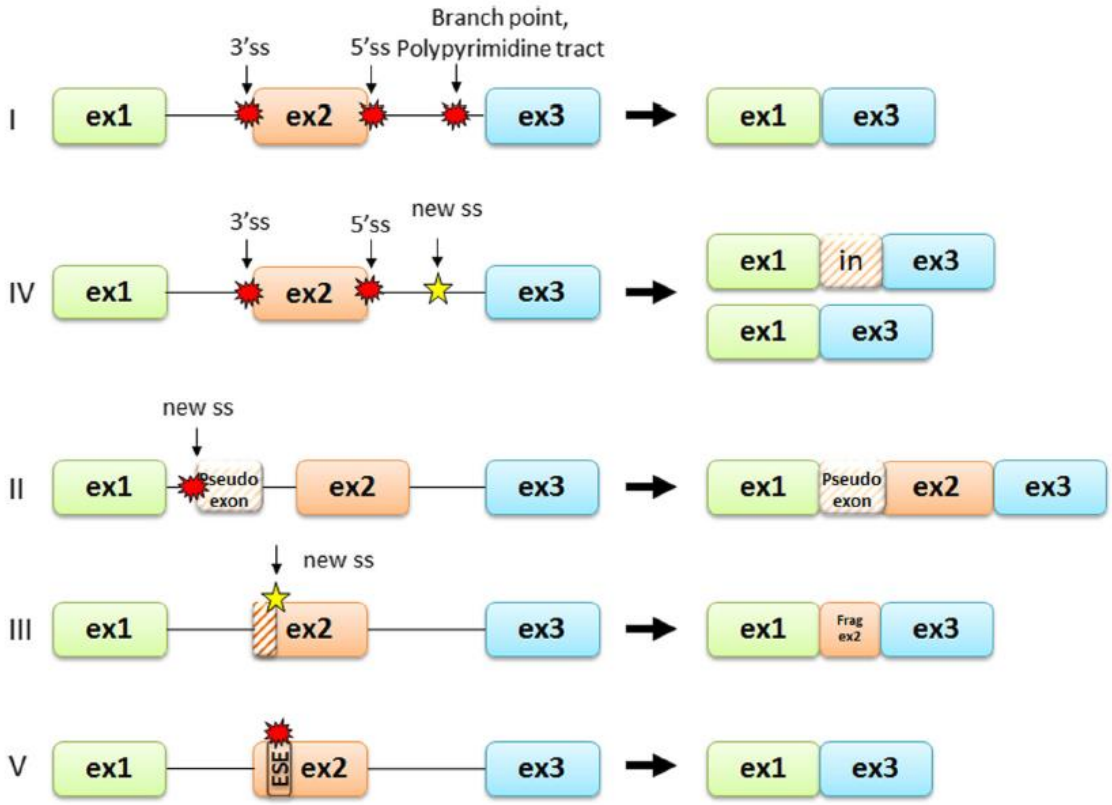

***Figura 4**. Tipos de mutaciones de splicing y los efectos que producen en el mRNA (Abramovicz & Gos, 2018).*

Las **mutaciones de tipo I y IV** se caracterizan por ser mutaciones en los sitios aceptores (*acceptor*) y donantes (*donor*) canónicos que afectan a secuencias fuertemente conservadas que definen los límites exón-intrón, imprescindibles para el buen funcionamiento del *splicing*. El sitio de *splicing* 5' (5'SS) (que tiene como secuencia conservada CAG/GUAAGU) y el sitio de *splicing* 3' (3'SS) (cuya secuencia suele ser NYAG/G) son reconocidos por unos elementos concretos del *spliceosome*: el snRNP de la proteína U1 reconoce y se une a la secuencia AG-GU complementaria en 5'SS y la proteína SF1 es reconocida y se une por la proteína U2AF65 al 3'SS (Fredericks *et al.*, 2015; Tazi *et al.*, 2009). Por tanto, cualquier variante en estas secuencias canónicas podría alterar la interacción entre el pre-mRNA y las proteínas implicadas en la eliminación del intrón. Las mutaciones más clásicas afectan a los residuos +1 y +2 en el sitio de *splicing donor* (5'SS) y a las posiciones -1 y -2 en el sitio de *splicing acceptor* (3'SS). Las mutaciones en estas regiones, generalmente, conducen a la omisión de un único exón completo (**tipo I**). Sin embargo, si el sitio de *splicing* es débil y la presencia de la mutación activa un sitio de *splicing* críptico en el exón o en un intrón vecino, el sitio alternativo se puede utilizar en el proceso de *splicing* (**tipo IV**), conduciendo a la inclusión de un fragmento de intrón o a la eliminación de un fragmento de exón, si el sitio de *splicing* críptico está presente en el intrón o en el exón, respectivamente (Wimmer, et al., 2007). En esta situación, se pueden generar varios transcritos distintos como en el caso de la variante c.1525-1G>A en el intrón 9 del gen *CFTR*. En este gen se observó la presencia de tres isoformas de mRNA diferentes que utilizaban distintos sitios de *splicing* alternativos ubicados dentro del intrón 10 y el exón

10 y carecían del exón 10 completo o algunos fragmentos, respectivamente (Ramalho *et al.*, 2003). En determinadas circunstancias, se pueden eliminar varios exones como en el caso de la variante c.1845+1G>A en el gen *NF1*, que conduce no solo al *skipping* del exón 16, sino también a la eliminación del exón 15 (Fang *et al.*, 2001).

Las **mutaciones de tipo II** son variantes intrónicas profundas, por lo general, sustituciones localizadas dentro de intrones grandes que dan como resultado la inclusión de un fragmento de intrón, denominado exón críptico o pseudoexón, en el transcrito maduro. Estas variantes crean nuevos sitios *acceptor* o *donor* que son reconocidos por el *spliceosome* y se usan en combinación con los sitios crípticos de *splicing* intrónicos existentes que no se habían empleado en el desarrollo normal de *splicing* (Wimmer, et al., 2007). También es posible que mutaciones intrónicas profundas den como resultado la creación de nuevos elementos reguladores (por ejemplo, potenciadores de *splicing*) y el reconocimiento de secuencias intrónicas específicas como si fueran secuencias exónicas (Vaz-Drago *et al.*, 2017). Las mutaciones intrónicas profundas no son comunes, pero su efecto sobre el *splicing* y la síntesis de proteínas es bastante significativo. Un ejemplo es la variante c.3718-2477C>T, que es una de las mutaciones más frecuentes en el gen *CFTR*, responsable de la fibrosis quística en la población polaca (Sobczyńska-Tomaszewska *et al.*, 2013). Esta mutación está ubicada dentro del intrón 19 y crea un nuevo sitio *donor* que da como resultado la inclusión de un exón críptico de 84 pb en el mRNA maduro. Este exón críptico contiene un codón de STOP dentro del ORF y, por lo tanto, la proteína traducida es más corta y no funcional (Sanz *et al.*, 2017).

Las **mutaciones de tipo III y V** son cambios que se encuentran dentro de las regiones codificantes del DNA, es decir, en los exones. Tales mutaciones exónicas podrían tener dos posibles efectos. En primer lugar, pueden introducir un nuevo sitio de *splicing* 5′ o 3′ o activar un sitio críptico que sería más fuerte que el original, lo que provocaría cambios en el procesamiento del pre-mRNA y la pérdida de un fragmento de exón (**tipo III**). En segundo lugar, la presencia de cambios exónicos que provocan la interrupción de los potenciadores del *splicing* exónico también pueden conducir a la omisión del exón completo (**tipo V**) (Wimmer, et al., 2007). Las mutaciones exónicas que causan alteraciones de *splicing* pueden clasificarse erróneamente como sinónimas, *missense* o *nonsense*. Por lo general, la presencia de tales variantes da como resultado la generación de dos transcritos diferentes a partir de un alelo mutado: uno tiene la longitud adecuada con el nucleótido modificado (dando un cambio aminoacídico o no dependiendo de la posición en el codón y del cambio de nucleótido), y el otro es más corto y carece del exón completo o un fragmento, debido a la actividad inespecífica del *spliceosome* (Nissim-Rafinia & Kerem, 2002). Como ejemplo, la presencia de la variante c.3362A>G en el gen *NF1* da como resultado dos isoformas del mRNA: una con el *splicing* correcto que contiene la sustitución que puede provocar un cambio *missense* erróneo a nivel de proteína (p.Glu1121Gly) y la otra carece del exón 20 completo (Xu *et al.*, 2014).

Esto indicaba que se pueden dividir las mutaciones puntuales en tres tipos distintos, según su posición en el DNA:

- Mutaciones intrónicas cercanas: cambios nucleotídicos que se encuentran en las posiciones adyacentes a los exones, aguas arriba o aguas abajo (*upstream* o *downstream*), que compondrían los tipos I y IV.
- Mutaciones intrónicas profundas: cambios nucleotídicos que se encuentran en el interior de los intrones alejadas de la región exónica, que compondrían el tipo II.
- Mutaciones exónicas: cambios nucleotídicos que se encuentran en las regiones codificantes del DNA, que corresponderían a los tipos III y V.

Con esta información, a partir del archivo descargado de *COSMIC*, se generó un *script* en *Python* que selecciona el gen, el transcrito, la posición cromosómica y la anotación de la mutación en el fichero. A partir de la anotación, para cada una de las líneas del fichero, primeramente, se observó si se trataba de un cambio puntual (es decir, que tuviera en su anotación el símbolo '>'). A continuación, se buscó si no tiene ningún símbolo '+' o '-', por lo que se tratará de una mutación exónica. En el caso de que tuviera estos símbolos, si el número a continuación de este era menor a 6, se indicaba como mutación intrónica cercana y, si era mayor, como mutación intrónica profunda. Se estableció el 6 como criterio para diferenciar entre las mutaciones intrónicas profundas y cercanas porque, en bibliografía, los cambios del tipo I y IV se producían como máximo hasta la sexta posición aguas arriba o aguas abajo del exón (Ibrahim *et al.*, 2007; Axelrod *et al.*, 2011), coincidiendo con las secuencias consenso de los sitios *donor* y *acceptor*.

En la **figura 5** se muestra cada uno de los tipos de anotaciones.

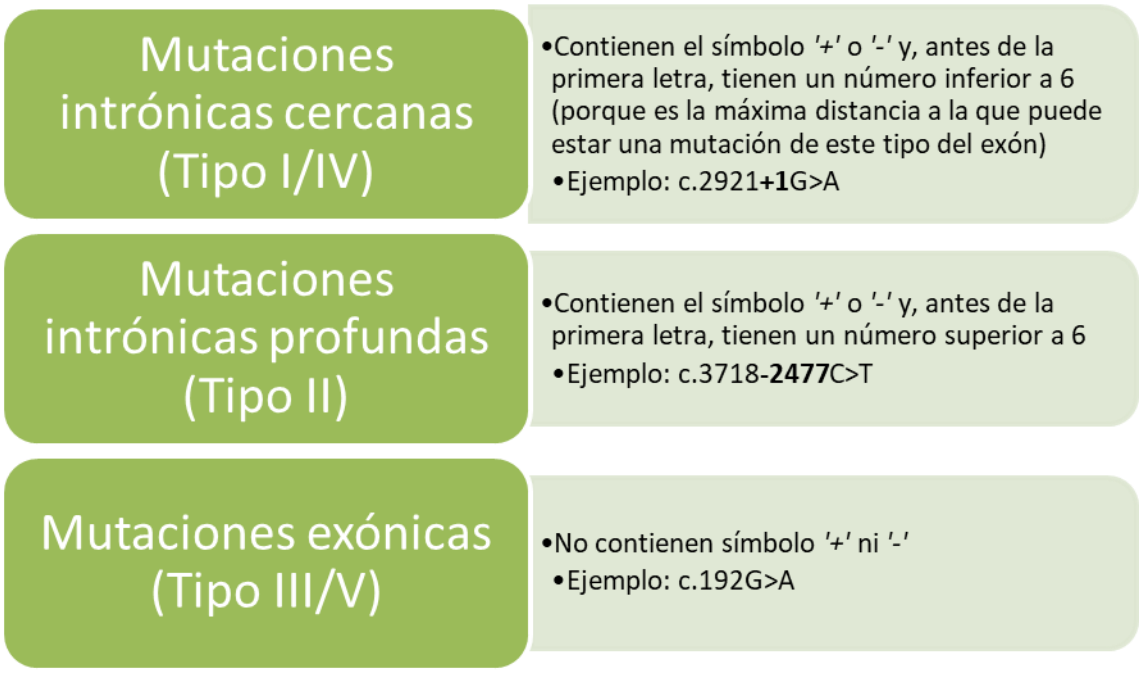

Figura 5. *Ejemplos de los tipos de mutaciones y criterio para diferenciarlos entre los grupos.*

Según los criterios mencionados, las mutaciones se separaron en tres ficheros distintos. Se generó un fichero a parte para aquellas mutaciones que tuvieran

una anotación que podría indicar un cambio puntual, pero daban error. El *script* completo se encuentra en el **Anexo I**.

2.1.2.2. Selección de los predictores a utilizar

A partir de una revisión bibliográfica, se seleccionaron diferentes predictores *in silico* para estudiar el efecto que puede tener una mutación en el *splicing*. Estas herramientas pueden ser de diferentes tipos, desde predictores con interfaces web hasta programas descargables que se puedan ejecutar en el ordenador. La **Tabla 1** refleja las diferentes características que tienen cada uno de los predictores encontrados, como el *input* y el *output* de cada método, el tipo de resultado obtenido o la complejidad de uso.

Tabla 1. *Predictores in silico de la búsqueda bibliográfica, así como el input y output necesario para ser utilizados.*

Predictor (Referencia(s) bibliográfica(s))	Input	Output	Notas
NetGene2 (Tate *et al.,* 2019; Brunak *et al.*, 1991)	Secuencia (se debe hacer una búsqueda para la secuencia WT y otra para la secuencia mutada).	Puntos de la secuencia donde hay *exon-intron boundaries*. Se compara entre ambas dónde están las diferencias, lo que supondrá que hay nuevos exones/intrones.	
Splice Site Prediction by Neural Network (NNSplice) (Reese *et al.*, 1997)	Secuencia (se debe hacer una búsqueda para la secuencia WT y otra para la secuencia mutada).	Puntos de la secuencia donde hay *exon-intron boundaries*. Se compara entre ambas dónde están las diferencias, lo que supondrá que hay nuevos exones/intrones.	
SplicePredictor (Brendel *et al.*, 2004)			Se debe descargar y ejecutar desde el ordenador. Según el artículo citado, "La herramienta de predicción del sitio de *splicing* (SplicePredictor) se distribuye con el código GeneSeqer. Un servidor web *SplicePredictor* está disponible en http://bioinformatics.iastate.edu/cgi-bin/sp.cg". Sin embargo, está página no ha sido encontrada.

Splice port (Dogan *et al.*, 2007)	Secuencia en formato FASTA		Da error al introducir la secuencia
SpliceView (Shapiro & Senapathy, 1987; Rogozin & Milanesi, 1997)	Secuencia		No aparece ningún resultado
GENSCAN (Burge & Karlin, 1997)	Secuencia (se debe hacer una búsqueda para la secuencia WT y otra para la secuencia mutada)	Puntos de la secuencia donde hay *exon-intron boundaries*. Se compara entre ambas dónde están las diferencias, lo que supondrá que hay nuevos exones/intrones.	
GeneSplicer (Pertea *et al.*, 2001)	Secuencia		Da error al introducir la secuencia (*Not found*)
MaxEntScan (Yeo & Burge, 2004)	Solo se introduce la región donde puede haber 5' splice sites / 3’ splice sites	Puntuaciones por diferentes métodos de dicha región sobre la efectividad del *splicing*. Cuando mayor sea el valor, el *splicing* es más efectivo (Eng *et al.*, 2004). Se ha predicho que los exones con sitios de *splicing* débiles contienen elementos cis auxiliares más abundantes, como ESE e ISE (Faustino & Cooper, 2003), lo que tal	No indica el efecto que puede tener el cambio.

		vez permita una regulación aún más compleja del *splicing* alternativo.	
Spliceman (Lim *et al.*, 2011; Lim & Fairbrother, 2012)	Secuencia FASTA (indicar en la posición concreta el cambio producido). Ejemplo: agcta(a/c)gatcg	Puntuación de la L1 más alta para los hexámeros que se generan a partir de una matriz 11-mer de la secuencia dada.	Cuanto más alto sea el rango del percentil, más probable es que la mutación puntual interrumpa el *splicing* (Lim *et al.*, 2011; Lim & Fairbrother, 2012)
CRYP-SKIP (Divina *et al.*, 2009)	Secuencia de nucleótidos del exón mutado junto con las secuencias intrónicas flanqueantes.	Proporciona la probabilidad de activación críptica del sitio de *splicing* (1) en contraposición a la omisión del exón (0). También muestra los valores de las variables predictoras más importantes.	
SROOGLE (Schwartz *et al.*, 2009)			*Offine*
Human Splicing Finder (Desmet *et al.*, 2009)	Gen, Transcrito y Código mutación a nivel CDS	Efecto que puede producir en el *splicing* con puntuación	Necesita registro y tiene número limitado de búsquedas (100)
Alamut Visual Software			FREE TRIAL DE 30 DÍAS
Mutation Forecaster			Mutation Forecaster es una web que tiene diferentes herramientas para estudiar las mutaciones que puden producir *splicing*.

			Entre ellas se encuentran *Shannon pipeline* (Análisis de mutaciones a escala genómica para predecir variantes que afectan la expresión génica (*splicing* y transcripción), *ASSEDA* (examina las consecuencias de una única variante en el splicng de mRNA) y *Veridical* (compara las variantes de *splicing* a escala del genoma con datos de RNA-Seq.). Necesita registro para poder utilizarse.
SVM-BPfinder (Corvelo *et al.*, 2010)	Secuencia WT región 3' en formato FASTA Seleccionar especie	Puntuación del *branch point* calculado con un clasificador SVM	No indica el efecto
IntSplice (Shibata *et al.*, 2016)	Insertar cromosoma y coordenada de la mutación	Indica si el *splicing* será normal o anormal	No indica el efecto
Variant Effect Predictor tool (McLaren *et al.*, 2010)	Coordenadas de las variantes y el cambio nucleotídico (bien con el código del gen o el identificador de la variante; entre otros formatos)	Indica el efecto de la variable	
Alamute Batch software			FREE TRIAL DE 30 DÍAS
ESEfinder (Cartegni *et al.*, 2003; Smith *et al.*, 2006)	Secuencia exónica (búsqueda para la	Mejor hit o todos los encontrados (Solo busca	Demasiada información, poco práctico

	secuencia WT y otra para la secuencia mutada a la vez).	ESE).	
RESCUE-ESE programs (Fairbrother *et al.*, 2002)	Secuencia exónica		*Not found*
HEXplorer score (Erkelenz *et al.*, 2014)	Secuencia RNA		Hay que guardar las secuencias en un fichero
ESRsearch (Fairbrother*et al.*, 2002; Goren *et al.*, 2006; Zhang & Chasin, 2004)	Secuencia exónica		Necesita registro
FAS-ESS (Wang *et al.*, 2004)	Secuencia exónica		*Not found*
SpliceAid2 (Piva *et al.*, 2012)	Secuencia exónica		*Internal Server Error*
SPANR tool (Xiong *et al.*, 2015)	Cromosoma, posición, identificador, alelo referencia, alelo mutado		Se envía la información pero nunca se obtienen resultados.
EX-SKIP (Raponi *et al.*, 2011)	Secuencias exónicas, dos alelos en formato FASTA	Tabla con resultados de sus análisis e indicación de qué efecto tendrá en *splicing*	Sólo analiza mutaciones exónicas.
HOT-SKIP (Raponi *et al.*, 2011)	Secuencias exónicas flanqueadas por 6 y 4 bp de secuencia intrónica en formato FASTA	Tabla que enumera las secuencias reguladoras de *splicing* predichas para todas las posibles mutaciones puntuales	Sólo analiza mutaciones exónicas.

Splicing Sequences Finder Splice site analysis	Secuencia o indicar gen		Not found (para ambos casos)
Programas GitHub https://github.com/friend1ws/SAVNet https://github.com/raphaelleman/SPiP			Deben ser instalados y son tediosos para ejecutar

2.1.2.2.1 Análisis de viabilidad de las herramientas

Para ver el resultado obtenido por cada predictor con cada uno de los tipos de mutaciones, se estudió la predicción obtenida con cada una de estas herramientas en mutaciones en las que se conocía el efecto que tienen en el *splicing.* En la **Tabla 2**, se muestran las mutaciones conocidas que fueron seleccionadas, así como su efecto real.

Tabla 2. *Mutaciones conocidas utilizadas para el análisis de viabilidad.*

Cambio	Tipo de mutación	Efecto real en el *splicing*	Referencia bibliográfica
ATM **c.2921+1G>A**	Mutación intrónica cercana	Pérdida exón 19 (efecto tipo I)	Gilad, et al., 1996
NF1 **c.1841+1G>A**	Mutación intrónica cercana	Pérdida exón 15 y 16 (efecto tipo I)	Fang *et al.*, 2001
COL5A1 **c.925-2A>G**	Mutación intrónica cercana	Dos transcritos: pérdida exón 6 y 7 / pérdida exón 7 (efecto tipo I)	Symoens *et al.*, 2011
OXCT1 **c.1248+5G>A**	Mutación intrónica cercana	Pérdida exón 12 y 13 (efecto tipo I)	Hori *et al.,* 2013
NF1 **c.288+1137C>T**	Mutación intrónica profunda	Inclusión exón críptico (118 pb) (efecto tipo II)	Svaasand *et al.*, 2015
CFTR **c.3718-2477C>T**	Mutación intrónica profunda	Inclusión exón críptico (84 pb) (efecto tipo II)	Sanz *et al.*, 2017
AR **c.2450-118A>G**	Mutación intrónica profunda	Dos transcritos: inclusión exón críptico (85 pb) e inclusión exón críptico (202 pb) (efecto tipo II)	Känsäkoski *et al.*, 2016
GLA **c.639+919G>A**	Mutación intrónica profunda	Inclusión exón críptico (57 pb) (efecto tipo II)	Palhais, et al., 2016
COL2A1 **c.192G>A**	Mutación exónica	Pérdida exón 2 (efecto tipo V)	McAlinden *et al.*, 2008
ACADM **c.382C>T**	Mutación exónica	Pérdida exón 5 (efecto tipo V)	Ward & Cooper, 2010
NF1 **c.3362A>G**	Mutación exónica	Dos transcritos: cambio missense / Pérdida exón 20 (efecto tipo V)	Xu *et al.*, 2014
BRCA1 **c.4484G>T**	Mutación exónica	Pérdida exón 14 (efecto tipo V)	Colombo *et al.*, 2013

Para poder realizar la búsqueda del efecto que tienen estas mutaciones, como se ha visto en la **Tabla 1**, en muchos casos es necesario tener como *input* la secuencia del gen en la región adyacente a la posición de la mutación. Esta secuencia fue obtenida a partir de *Ensembl* (Zerbino, et al., 2018). Si se tenía información sobre en qué transcrito se encuentra el cambio, se obtuvo dicha secuencia y, si no, la del transcrito mayoritario. A partir de aquí, se buscaron los cambios que se producían entre la secuencia *wild type* (WT) o canónica, la

cual contenía el nucleótido de referencia; y la secuencia mutante, la cual contenía el nucleótido alternativo o al que había mutado el cambio de referencia; a partir de los diferentes predictores seleccionados.

A continuación, para cada predictor se llevó a cabo un análisis de concordancia con respecto al efecto real de la mutación. Se obtuvo, también, para cada uno, la tabla de confusión correspondiente. Esto nos permitió reducir el número de predictores empleados en los siguientes pasos.

2.1.2.3. Análisis comparativo de predictores de *splicing*

2.1.2.3.1 Selección de variantes

Dada la limitación de pruebas de uno de los predictores (HSF: Tabla 1), se seleccionaron de manera aleatoria, a partir de un *script* de *Python*, 29 variantes de cada uno de los tres tipos de variantes obtenidos a partir de la base de datos (mutaciones intrónicas cercanas, intrónicas profundas y exónicas; referencia a la sección 3.2.1). El *script* para obtener estas variantes se encuentra en el **Anexo II**.

Una vez seleccionadas, se obtuvo la secuencia correspondiente al exón más próximo a la mutación, así como parte de la secuencia intrónica adyacente a este, a partir de la base de datos de *Ensembl* (Zerbino, et al., 2018) en la versión 38 del genoma humano (GRCh38.p13). Se empleó esta versión dado que los datos del fichero estaban anotados en esta versión. A partir de aquí, se anotaron los cambios que se observaron entre las predicciones para la secuencia *wild type* (WT) o canónica y la secuencia mutante, en algunos casos, o el resultado del efecto de la mutación, en otros predictores.

2.1.2.3.2 Comparación de métodos

Para comparar la eficacia de los métodos es importante determinar si el método es capaz de identificar *splicing* cuando lo hay y no detectarlo cuando no lo hay. Para ello hay que poder definir los valores que nos van a determinar la eficiencia de un método diagnóstico.

- Un verdadero positivo (*True Positive*) es un resultado en el que el modelo predice correctamente la clase positiva.
- Un verdadero negativo (*True Negative*) es un resultado en el que el modelo predice correctamente la clase negativa.
- Un falso positivo (*False Positive)* es un resultado en el que el modelo predice incorrectamente la clase positiva
- Un falso negativo (*False Negative*) es un resultado en el que el modelo predice incorrectamente la clase negativa.

Todos estos valores se obtienen en la denominada como tabla de confusión. Con estos valores se pueden obtener una serie de parámetros:

- Sensibilidad: proporción de ejemplos positivos que son correctamente clasificados
- Especificidad: proporción de ejemplos negativos que son correctamente clasificados
- Precisión: proporción de ejemplos positivos que son realmente positivos
- Exhaustividad: número de verdaderos positivos frente al total de positivos) de cada uno de los predictores. Su valor es el mismo al de la sensibilidad.

Como la mayoría de las veces no se conoce de antemano dicho efecto, en base a los resultados del apartado anterior (sección 2.2.2.2.1) se optó por tomar el HSF como método de referencia para determinar el efecto real, al tratarse del método que más se aproximó a ello en el análisis de viabilidad.

El análisis estadístico para llevar a cabo la comparación de métodos consistió en primer lugar en comparar descriptivamente los resultados obtenidos por cada método, incluido el método HSF, que en este caso actúa como referencia. En segundo lugar, se crearon tablas de confusión para cada método. La tabla de conusión recoge los verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN). Finalmente, cuando fue posible se construyó una curva ROC y se calculó el área bajo la curva (*Area Under the* Curve, o por sus siglas en inglés, AUC). Una curva ROC (*Receiver Operating Characteristic*) es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la proporción de verdaderos positivos frente a la proporción de falsos positivos (Hoo *et al.*, 2017).

2.2 Resultados

2.2.1. Selección y estudio de base de datos

Para obtener los datos del Catálogo de Mutaciones Somáticas en Cáncer (*COSMIC*) fue necesario acceder al portal web correspondiente. En este portal, existen diferentes archivos en los que se encuentran anotadas las mutaciones. Se seleccionó el fichero constando de anotaciones HGSV. Se trata del archivo *CosmicMutantExport.tsv.gz*. Se observó que el archivo constaba de 47.222.280 variantes con diferentes tipos de atributos. La anotación a nivel de CDS se extrajo del fichero. A partir de esta información, se seleccionaron las 43.585.350 mutaciones comportando un cambio de nucleótidos (caracterizadas porque tienen el símbolo '>') o no. De estas mutaciones, hay 18.296.779 que son exónicas (cambios puntuales que no tienen ni el símbolo '+' ni '-'), 10.782.527 intrónicas *downstream* (cambios puntuales que tienene símbolo '+') y 10.565.679 intrónicas *upstream* (cambios puntuales que tienene símbolo '+'). La diferencia hasta 43.585.350 (13.398.338 líneas) son cambios que están relacionados con la *5'UTR* como c.-59A>C, c.-517+200G>A o c.-516-964G>A.

Además, también hay 1.101.775 duplicaciones (tienen la identificación *dup* en su anotación), 1.147.480 deleciones (tienen la identificación *del* en su anotación), 278.969 inserciones (tienen la identificación *ins* en su anotación), 60.671 inserciones/deleciones (tienen la identificación *ins ins* y *del* en su anotación) y 1.046.035 de otros tipos. La **figura 6** muestra estos resultados de manera esquemática.

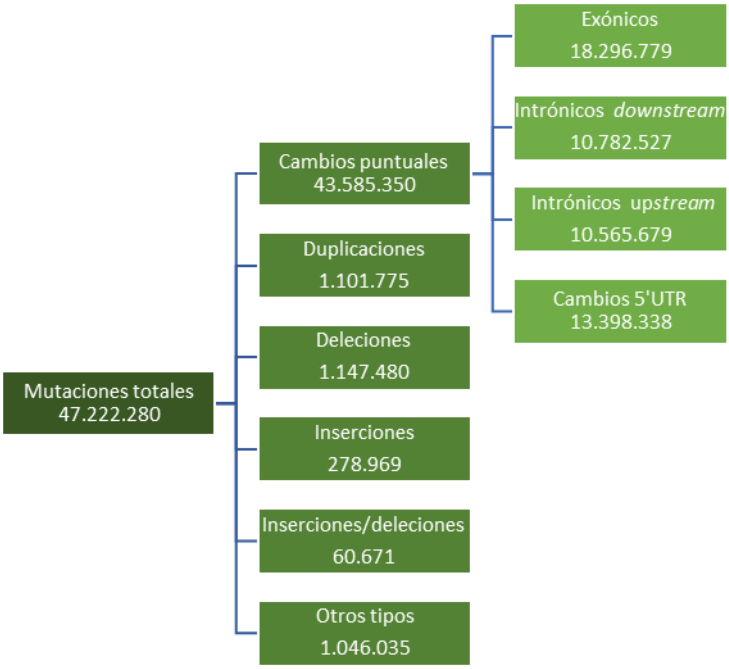

Figura 6. *Distribución del tipo de mutaciones en la base de datos COSMIC*.

2.2.2 Detección de variantes genéticas que afectan al *splicing*

2.2.2.1 Caracterización de variables

Una vez identificadas las características de cada una de las variantes (sección 2.1.2.1.), se ejecutó un s*cript de Python* (**Anexo I**) para separar las mutaciones puntuales en tres archivos distintos. El primer archivo contiene 436.982 variantes intrónicas cercanas (variantes de tipo I o IV). El segundo archivo contiene 20.911.221 variantes intrónicas profundas (variantes de tipo II). Finalmente, el tercer archivo contiene 18.296.160 variantes exónicas (variantes de tipo III o V). Se generó un fichero adicional para guardar las mutaciones puntuales con alguna nomenclatura extraña.

2.2.2.2. Selección de los predictores a utilizar

De los predictores de la **Tabla 1**, las herramientas que están en un color más claro se descartaron automáticamente por varias posibles razones:

- Se trataba de servidores de pago (*Alamute Batch software, Alamut Visual Software, MutationForecaster*, *ESRsearch*).
- No se encontraba operativa (*SROOGLE*)

- Al intentar introducir la secuencia, no da una página con resultados (*RESCUE-ESE programs, FAS-ESS, SpliceAid2, SPANR tool, Splicing Sequences Finder Splice site analysis*)
- Eran tediosos y complicados de utilizar (Programas GitHub)

El resto pasaron a emplearse en el punto siguiente para observar las predicciones obtenidas para cuyo el efecto en el *splicing* era conocido (**Tabla 2**).

La lista final de predictores a testar es:

- *NetGene2*
- *NNSplice*
- *GENSCAN*
- *MaxEntScan*
- *Spliceman*
- *CRYP-SKIP*
- *Human Splicing Finder*
- *SVM-BPfinder*
- *IntSplice*
- *Variant Effect Predictor tool*
- *ESEfinder*
- *EX-SKIP*
- *HOT-SKIP*

2.2.2.2.1 Análisis de viabilidad de las herramientas

Se ejecutó cada uno de los predictores anteriores a las variantes descritas en la **tabla 2**. Los resultados obtenidos se muestran en las **tablas 3** a **5**. Los resultados completos se encuentran en el **Anexo III**.

Tabla 3. *Resultados para los cambios conocidos de tipo intrónico cercano (tipo I o IV) generados por los predictores.*

Efecto para:	***ATM* c.2921+1G>A**	***NF1* c.1845+1G>A**	***COL5A1* c.925-2A>G**	***OXCT1* c.1248+5G>A**
Efecto real	**Pérdida exón 19**	**Pérdida exón 15 y 16**	**Dos transcritos: pérdida exón 6 y 7 / pérdida exón 7**	**Pérdida exón 12 y 13**
NetGene2	No efecto	No efecto	No efecto	Pérdida sitio *donor*, *exon skipping*
NNSplice	Pérdida sitio *donor* exón 19	Pérdida sitio *donor* exón 15	Pérdida sitio aceptor exón 7, *exon skipping*	Pérdida sitio *donor*, *exon skipping*
GENSCAN	No resultados	No resultados	No resultados	No resultados
MaxEntScan	No efecto	No efecto	No efecto	No efecto
Spliceman	Afectando al *splicing* (68%)	Afectando al *splicing* (68%)	Afectando al *splicing* (83%)	Afectando al *splicing* (76%)
CRYP-SKIP	No efecto (no útil para mutaciones intrónicas)	No efecto (no útil para mutaciones intrónicas)	No efecto (no útil para mutaciones intrónicas)	No efecto (no útil para mutaciones intrónicas)
Human Splicing Finder	Uso nuevo sitio *donor*, probable pérdida de un exón	Alteración sitio *donor*, *exon skipping*	Pérdida sitio aceptor	Alteración sitio *donor*, *exon skipping*
SVM-BPfinder	Uso nuevo sitio de *splicing*, probable pérdida de un exón	No efecto	No efecto	Uso nuevo sitio de *splicing*, probable pérdida de un exón
IntSplice	No resultados	No resultados	No resultados	No resultados
Variant Effect Predictor tool	Variante sitio *donor*, posible pérdida de un exón	Alteración sitio *donor*, *exon skipping*, y activación NMD	Alteración sitio aceptor, *exon skipping*	Uso nuevo sitio *donor*, posible pérdida de un exón
ESEfinder	Pérdida sitio *donor*	Pérdida sitio *donor*	Pérdida sitio aceptor	Pérdida sitio *donor*
EX-SKIP	No resultados	*Exon skipping*	No efecto	No efecto
HOT-SKIP	No efecto	No efecto	No efecto	No efecto

Tabla 4. *Resultados para los cambios conocidos de tipo intrónico profundo (tipo II) generados por los predictores.*

Efecto para…	*NF1* c.288+1137G>A	*CFTR* c.3718-2477C>T	*AR* c.2450-118A>G	*GLA* c.639+919G>A
Efecto real	**Inclusión exón críptico (118 pb)**	**Inclusión exón críptico (84 pb)**	**Dos transcritos: inclusión exón críptico (85 pb) e inclusión exón críptico (202 pb)**	**Inclusión exón críptico (57 pb) dentro del exón (in-frame)**
NetGene2	Nuevo sitio *donor,* posible inclusión exón críptico (870 pb)	Nuevo sitio *donor,* posible inclusión exón críptico	Nuevo sitio *donor,* posible inclusión exón críptico (57 pb)	Nuevo sitio *donor,* posible inclusión exón críptico (57 pb)
NNSplice	Nuevo sitio *donor,* posible inclusión exón críptico (120 pb)	Nuevo sitio *donor,* posible inclusión exón críptico	Nuevo sitio *donor,* 2 posibles inclusiones de exón críptico (46 u 87 pb)	No efecto
GENSCAN	No resultados	No resultados	No resultados	No resultados
MaxEntScan	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)
Spliceman	Afectando al *splicing* (66%)	Afectando al *splicing* (65%)	Afectando al *splicing* (73%)	Afectando al *splicing* (61%)
CRYP-SKIP	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)
Human Splicing Finder	Activación sitio *donor* críptico, inclusión exón críptico	Alteración ESE/ESS, posible alteración *splicing*	Activación sitio *aceptor*. Potencial alteración *splicing*	Alteración ESE/ESS y/o activación sitio aceptor, potencial alteración *splicing*
SVM-BPfinder	No efecto	No efecto	No efecto	Aparición nuevo BP en la secuencia mutante, posible alteración del *splicing.*
IntSplice	No resultados	No resultados	No resultados	No resultados
Variant Effect Predictor tool	Efecto en el *splicing*, produce NMD	Efecto en el *splicing*, produce NMD	Efecto en el *splicing*, produce NMD	Efecto en el *splicing*, produce NMD y es una variante *downstream* que afecta al 3'UTR
ESEfinder	Nuevo 5'SS con mayor probabilidad que nuevo 3'SS. Efecto en el *splicing*	Nuevo 5'SS más fuerte. Efecto en el *splicing*	Nuevo 5'SS más fuerte. Efecto en el *splicing*	Pérdida 3'SS. Efecto en el *splicing*
EX-SKIP	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)
HOT-SKIP	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)	No resultados (predictor no útil para variantes intrónicas profundas)

Tabla 5. *Resultados para los cambios conocidos de tipo exónico (tipo III o V) generados por los predictores.*

Predictor	*COL2A1* c.192G>A	*ACADM* c.382C>T	*NF1* c.3362A>G	*BRCA1* c.4484G>T
Efecto real	**Pérdida exón 2**	**Pérdida exón 5**	**Missense / Pérdida exón 20**	**Pérdida exón 14**
NetGene2	No efecto	No efecto	No efecto	No efecto
NNSplice	No efecto	No efecto	No efecto	No efecto
GENSCAN	No resultados	No resultados	No resultados	No resultados
MaxEntScan	No efecto	No efecto	No efecto	No efecto
Spliceman	Afectando al *splicing* (72%)	Afectando al *splicing* (50%)	Afectando al *splicing* (49%)	Afectando al *splicing* (78%)
CRYP-SKIP	No efecto	No efecto	No efecto	No efecto
Human Splicing Finder	Disrupción ESE/ESS, probable *exon skipping*	No efecto	Activación sitio *donor* críptico, potencial alteración *splicing*	Alteración ESE/ESS
SVM-BPfinder	No resultados	No efecto	No efecto	Cambio región 5'SS posible efecto
IntSplice	No resultados	No resultados	No resultados	No resultados
Variant Effect Predictor tool	*Non coding transcript exon variant,* produce NMD afectando al *splicing*	*Non coding transcript exon variant,* produce NMD afectando al *splicing,* y *missense variant*	*Non coding transcript exon variant,* produce NMD afectando al *splicing,* y *missense variant*	Variante en región de *splicing,* afecta el 3'UTR y, produce NMD afectando al *splicing*
ESEfinder	Modificación ESE interno, cambios en el *splicing* Pérdida 5'SS	Modificación ESE interno, cambios en el *splicing* Generación 3'SS	Modificación ESE interno, cambios en el *splicing* SS sin conclusiones	Modificación ESE interno, cambios en el *splicing* 5'SS más débil
EX-SKIP	Probabilidad de *exon skipping*	Probabilidad de *exon skipping*	Probabilidad de *exon skipping*	No efecto
HOT-SKIP	No efecto	No efecto	Afectando al *splicing* (0.25)	No efecto

A continuación, se llevó a cabo el análisis de concordancia para cada uno de los predictores. Primeramente, se estudió si el predictor producía algún tipo de efecto o no. También se estudió si la predicción coincidía o no con el efecto real. Los resultados se encuentran en el **Anexo IV**. El porcentaje de veces en que hay efecto para cada uno de los predictores, así como el porcentaje de veces en que en la obtención del efecto real en la predicción, distinguiendo entre todas y por tipos de mutación, se muestra en la **Tabla 6**.

Tabla 6. *Porcentaje de veces en que hay efecto y porcentaje de veces que la predicción coincide con el efecto real para cada predictor, distinguiendo entre todas las mutaciones y por cada tipo.*

	Porcentaje presencia efecto / Porcentaje acierto en efecto			
Predictor	**Todas las mutaciones**	**Mutaciones intrónicas cercanas**	**Mutaciones intrónicas profundas**	**Mutaciones exónicas**
NetGene2	41.70% / 41.70%	25.00% / 25.00%	100.00% / 100.00%	0.00% / 0.00%
NNSplice	58.30% / 58.3%	100.00% / 100.00%	75.00% / 75.00%	0.00% / 0.00%
Spliceman	83.30% / 0.00%	100% / 0.00%	100.00% / 0.00%	50.00% / 0.00%
Human Splicing Finder	91.70% / 66.70%	100.00% / 100.00%	100.00% / 75.00%	75.00% / 25.00%
SVM-BPfinder	33.30% / 33.30%	50.00% / 50.00%	25.00% / 25.00%	25.00% / 25.00%
Variant Effect Predictor tool	100.00% / 41.70%	100.00% / 100.00%	100.00% / 0.00%	100.00% / 25.00%
ESEfinder	100.00% / 66.70%	100.00% / 100.00%	100.00% / 75.00%	100.00% / 25.00%
EX-SKIP	33.30% / 33.30%	25.00% / 25.00%	0.00% / 0.00%	75.00% / 75.00%
HOT-SKIP	8.30% / 8.30%	0.00% / 0.00%	0.00% / 0.00%	25.00% / 25.00%

Seguidamente, se muestran los resultados de la tabla de confusión para el estudio de la presencia o no de efecto (**Tabla 7**) y para la concordancia de la predicción con el efecto real (**Tabla 8**), así como los estimadores para cada uno de los predictores. Como solo se tienen ejemplos de la presencia de efecto y no de la ausencia de este, en la tabla no se obtienen ni verdaderos negativos ni falsos positivos y, por tanto, tampoco se puede calcular la especificidad. Además, en este caso en que solo hay verdaderos positivos y falsos negativos, la sensibilidad y la exhaustividad serán igual al porcentaje de veces que coinciden predicción y referencia y la precisión es del 100% en todos los casos, ya que no hay falsos positivos.

Tabla 7. *Resultados de la tabla de confusión para el estudio de la presencia o no de efecto.*

Predictor	Verdaderos Positivos	Falsos Negativos	Sensibilidad	Precisión	Exhaustividad
NetGene2	5	7	41.70%	100.00%	41.70%
NNSplice	7	5	58.30%	100.00%	58.30%
Spliceman	10	2	83.30%	100.00%	83.30%
Human Splicing Finder	11	1	91.70%	100.00%	91.70%
SVM-BPfinder	4	8	33.30%	100.00%	33.30%
Variant Effect Predictor tool	12	0	100.00%	100.00%	100.00%
ESEfinder	12	0	100.00%	100.00%	100.00%
EX-SKIP	4	8	33.30%	100.00%	33.30%
HOT-SKIP	1	11	8.30%	100.00%	8.30%

Tabla 8. *Resultados de la tabla de confusión para la comprobación si la predicción obtiene el efecto real o no.*

Predictor	Verdaderos Positivos	Falsos Negativos	Sensibilidad	Precisión	Exhaustividad
NetGene2	5	7	41.70%	100.00%	41.70%
NNSplice	7	5	58.30%	100.00%	58.30%
Spliceman	0	12	0.00%	100.00%	0.00%
Human Splicing Finder	8	4	66.70%	100.00%	66.70%
SVM-BPfinder	4	8	33.30%	100.00%	33.30%
Variant Effect Predictor tool	5	7	41.70%	100.00%	41.70%
ESEfinder	8	4	66.70%	100.00%	66.70%
EX-SKIP	4	8	33.30%	100.00%	33.30%
HOT-SKIP	1	11	8.30%	100.00%	8.30%

2.2.2.3 Análisis comparativo de predictores de *splicing*

4.2.3.1 Selección de variantes

Con el *script* del **Anexo II**, se obtuvieron de manera aleatoria 29 variantes de cada uno los tres documentos que se habían separado anteriormente (sección 2.2.2.1). En el **Anexo V** se expone una lista de estas variantes seleccionadas, con la información del gen, transcrito, posición cromosómica (versión 38 del genoma humano) y su anotación HGSV.

4.2.3.2. Comparación de métodos

Según los resultados obtenidos en el apartado anterior, el listado final de métodos a comparar es:

- *NetGene2*
- *NNSplice*
- *Spliceman*
- *Human Splicing Finder*
- *SVM-BPfinder*
- *Variant Effect Predictor tool*
- *ESEfinder*
- *EX-SKIP (sólo para mutaciones exónicas)*
- *HOT-SKIP (sólo para mutaciones exónicas)*

Los resultados obtenidos para cada tipo de variantes se muestran, de manera resumida, en las **tablas 9** a **11**. Los resultados completos se encuentran en el **Anexo III**.

Tabla 9. *Resultados para los cambios de la base de datos de tipo intrónico cercano (tipo I o IV) generados por los predictores (parte 1 de 3).*

Predictor Cambio	Human Splicing Finder	NetGene2	NNSplice	Spliceman	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder
CACNA1C **c.1218-2A>G**	Sitio *acceptor* "roto"	No efecto	Pérdida sitio *acceptor*	Alteración en el *splicing* (78%)	No efecto	Variante afectando al sitio *acceptor*	Pérdida sitio *acceptor*
GIT1 **c.405+5G>C**	Sitio *donor* "roto". Activación sitio *acceptor*	No efecto	Pérdida sitio *donor*	Alteración en el *splicing* (52%)	No efecto	Variante afectando a una región de *splicing*	Sitio *donor* más débil
CHD4 **c.3820-1G>T**	No efecto	Pérdida sitio *acceptor* del exón	Pérdida sitio *acceptor*	Alteración en el *splicing* (84%)	No efecto	Variante afectando al sitio *acceptor*	Pérdida sitio *acceptor*
PIR **c.273+1G>A**	Alteración sitio *donor*	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (53%)	No efecto	Variante afectando al sitio *donor*	Pérdida sitio *donor*
UEVLD **c.821-1G>C**	Sitio *acceptor* "roto"	Pérdida sitio *acceptor*	No efecto	Alteración en el *splicing* (54%)	No efecto	Variante afectando al sitio *acceptor*	No efecto
CTCFL **c.755-1G>T**	Sitio *acceptor* "roto"	Pérdida sitio *acceptor*	Pérdida sitio *acceptor*	Alteración en el *splicing* (61%)	No efecto	Variante afectando al sitio *acceptor*	Pérdida sitio *acceptor*
NF2 **c.1488-1G>C**	Sitio *acceptor* "roto"Activación sitio *acceptor* crítico	Pérdida sitio *acceptor*	No efecto	Alteración en el *splicing* (64%)	No efecto	Variante afectando al sitio *acceptor*	Pérdida sitio *acceptor*
U2AF1 **c.44+1G>A**	Sitio *donor* "roto"	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (66%)	No efecto	Variante afectando al sitio *donor*	Sitio *donor* más débil
MAP4K4 **c.640-1G>T**	Sitio *acceptor* "roto". Activación sitio *acceptor* crítico	Pérdida sitio *acceptor*	Pérdida sitio *acceptor* y activación *acceptor* críptico	Alteración en el *splicing* (72%)	No efecto	Variante afectando al sitio *acceptor*	Pérdida sitio *acceptor*
RHCE **c.939+1G>T**	Sitio *donor* "roto"	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (58%)	No efecto	Variante afectando al sitio *donor*	Sitio *donor* más débil

Tabla 9. *Resultados para los cambios de la base de datos de tipo intrónico cercano (tipo I o IV) generados por los predictores (parte 2 de 3).*

Predictor Cambio	Human Splicing Finder	NetGene2	NNSplice	Spliceman	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder
TP53 **c.375+1G>A**	Sitio *donor* "roto"	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (52 %)	No efecto	Variante afectando sitio *donor*	Pérdida sitio *donor*
KLHL2 **c.1480+2T>C**	Sitio *donor* "roto"	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (60 %)	No efecto	Variante afectando sitio *donor*	Sitio *donor* debilitado
TSPAN17 **c.747+1G>T**	Sitio *donor* "roto"	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (64 %)	No efecto	Variante afectando sitio *donor*	Pérdida sitio *donor*
PAMR1 **c.714+5G>C**	Sitio *donor* "roto" Activación sitio *acceptor* críptico	No efecto	No efecto	Alteración en el *splicing* (51 %)	Activación BP, posible efecto en el *splicing*	Variante afectando región de *splicing*	Sitio *donor* más débil y *acceptor* ligeramente más fuerte
CACNA1B **c.2093-3C>A**	No efecto	No efecto	No efecto	Alteración en el *splicing* (42 %)	No efecto	Variante afectando región de *splicing*	Pérdida sitio *acceptor*
DIDO1 **c.2214+2T>C**	Sitio *donor* "roto"	Pérdida sitio *dono*	Pérdida sitio *donor*	Alteración en el *splicing* (84 %)	No efecto	Variante afectando sitio *donor*	Sitio *donor* debilitado
ARHGAP28 **c.955-1G>T**	Sitio *acceptor* "roto". Activación sitio *donor* críptico	Pérdida sitio *acceptor*	Pérdida sitio *acceptor*	Alteración en el *splicing* (90 %)	No efecto	Variante afectando sitio *acceptor*	Pérdida sitio *acceptor*
FRMPD4 **c.813+3C>G**	Activación sitio *donor* críptico	No efecto	No efecto	Alteración en el *splicing* (73 %)	BP más débil	Variante afectando región de *splicing*	Sitio *donor* modificado
NIN **c.1546-2A>G**	Sitio *acceptor* "roto". Activación sitio *donor* críptico	Pérdida sitio *acceptor*	Pérdida sitio *acceptor*	Alteración en el *splicing* (69 %)	No efecto	Variante afectando sitio *acceptor*	Sitio *donor* activado y *acceptor* ligeramente más débil o perdido
CLASRP **C.100-1G>A**	Sitio *acceptor* "roto"	Pérdida sitio *acceptor*	Pérdida sitio *acceptor*	Alteración en el *splicing* (78 %)	No efecto	Variante afectando sitio *acceptor*	Pérdida sitio *acceptor*

Tabla 9. *Resultados para los cambios de la base de datos de tipo intrónico cercano (tipo I o IV) generados por los predictores (parte 3 de 3).*

Predictor Cambio	Human Splicing Finder	NetGene2	NNSplice	Spliceman	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder
CHL1 **c.197+1G>A**	Alteración sitio *donor*	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (65 %)	No efecto	Variante sitio *donor* de *splicing*	Pérdida sitio *donor*
NPR3 **c.892+1G>T**	Alteración sitio *donor*. Activación *donor* críptico	Pérdida sitio *donor* y nuevo *acceptor*	Pérdida sitio *donor*	Alteración en el *splicing* (68 %)	BP más débil	Variante sitio *donor* de *splicing*	Pérdida sitio *donor*
MFSD8 **c.554-1G>T**	Alteración sitio *acceptor*	Nuevo *acceptor*	No efecto	Alteración en el *splicing* (92 %)	BP más fuerte	Variante sitio *acceptor* de *splicing*	Sitio *acceptor* más fuerte
SEMA4A **c.685+1G>A**	Alteración sitio *donor*	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (68 %)	Pérdida BP	Variante sitio *donor* de *splicing*	Pérdida sitio *donor*
GFPT2 **c.2004+4C>T**	Alteración sitio *donor*	Alteración sitio *donor*	Alteración sitio *donor*	Alteración en el *splicing* (81 %)	No efecto	Variante región de *splicing*, variante intrónica	Sitio *acceptor* más débil
PTPRT **c.1865+1G>T**	Alteración sitio *donor*	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (73 %)	No efecto	Variante sitio *donor* de *splicing*	Pérdida sitio *donor*
STK11 **c.465-1G>T**	Alteración sitio *acceptor*. Activación *acceptor* críptico	Pérdida sitio *acceptor*	No efecto	Alteración en el *splicing* (52 %)	No efecto	Variante sitio *acceptor* de *splicing*	Pérdida sitio *acceptor*
NXF2 **c.233+2T>C**	Alteración sitio *donor*	Pérdida sitio *donor*	Pérdida sitio *donor*	Alteración en el *splicing* (47 %)	No efecto	Variante sitio *donor* de *splicing*	Sitio *donor* más débil
GPX5 **c.242-3C>T**	No efecto	Pérdida sitio *acceptor*	No efecto	Alteración en el *splicing* (70 %)	No efecto	Variante región de *splicing* e intrónica	Activación sitio *donor*

Tabla 10. *Resultados para los cambios de la base de datos de tipo intrónico profundo (tipo II) generados por los predictores (parte 1 de 3).*

Predictor **Cambio**	**Human Splicing Finder**	**NetGene2**	**NNSplice**	**Spliceman**	**SVM-BPfinder**	**Variant Effect Predictor tool**	**ESEfinder**
ATXN2L **c.616+79G>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (60%)	BP más débil	Variante intrónica	Sitio *donor* muy débil
LZTR1 **c.1785+21A>G**	No efecto	No efecto	No efecto	Alteración en el *splicing* (53 %)	No efecto	Variante intrónica	No efecto
CES1 **c.1318+724G>C**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (53 %)	No efecto	Variante intrónica	Sitio *donor* ligeramente más fuerte
CELF4 **c.801+93C>A**	Activación sitio *donor* críptico	No efecto	No efecto	Alteración en el *splicing* (52 %)	No efecto	Variante intrónica	No efecto
DDR1 **c.418-287G>T**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (62 %)	No efecto	Variante intrónica	No efecto
KRT17 **c.433-195T>C**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (74 %)	No efecto	Variante intrónica	Sitio *donor* nuevo (poco probable)
DGKI **c.1403-841A>T**	Activación sitio *acceptor* críptico, activación exón críptico	Activación *acceptor* críptico	Activación *acceptor* críptico	Alteración en el *splicing* (70 %)	No efecto	Variante intrónica	Sitio *acceptor* más fuerte
GADD45GIP **c.350+644G>C**	No efecto	No efecto	No efecto	Alteración en el *splicing* (48 %)	No efecto	Variante intrónica	Sitio *donor* ligeramente más débil
CACNA1G **c.1924+645G>A**	Alteración ratio ESE/ESS	Activación *acceptor* críptico	No efecto	Alteración en el *splicing* (65%)	No efecto	Variante intrónica	Sitio *donor* ligeramente más débil
CALD1 **c.1171-428G>C**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (56 %)	BP más débil	Variante intrónica	No efecto

Tabla 10. *Resultados para los cambios de la base de datos de tipo intrónico profundo (tipo II) generados por los predictores (parte 2 de 3).*

Predictor Cambio	Human Splicing Finder	NetGene2	NNSplice	Spliceman	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder
OPRM1 **c.864+1405G>A**	Alteración ratio motivos ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (70 %)	Posible pérdida BP	Variante intrónica	No efecto
PTPRT **c.3771+1049T>G**	Alteración ratio motivos ESE/ESS	Activación *acceptor* críptico	No efecto	Alteración en el *splicing* (73 %)	No efecto	Variante intrónica	Sitio *acceptor* más débil/perdido
ELP4 **c.259+1970A>G**	Activación sitio *donor* críptico	No efecto	No efecto	Alteración en el *splicing* (66 %)	No efecto	Variante intrónica	Sitio *donor* ligeramente más fuerte
NEB **c.6915+1336A>G**	Alteración ratio motivos ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (74 %)	No efecto	Variante intrónica	No efecto
GABRB2 **c.353-1057G>T**	Activación sitio *donor* críptico	No efecto	No efecto	Alteración en el *splicing* (54 %)	No efecto	Variante intrónica	Activación sitio *donor*
COG6 **c.789-1074C>T**	No efecto	No efecto	Sitio *acceptor* más fuerte	Alteración en el *splicing* (67 %)	No efecto	Variante intrónica	No efecto
CPLANE1 **c.938+1393T>A**	No efecto	No efecto	No efecto	Alteración en el *splicing* (54 %)	No efecto	Variante intrónica	No efecto
ABCA8 **c.2765-1855T>G**	Alteración ratio motivos ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (78 %)	Activación BP	Variante intrónica	No efecto
AC093668.1 **c.425+1352A>G**	No efecto	Activación *acceptor* críptico	No efecto	Alteración en el *splicing* (82 %)	No efecto	Variante intrónica	No efecto
ATP5MF-PTCD1 **c.121+1825T>A**	Alteración ratio motivos ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (65 %)	Activación BP	Variante intrónica	No efecto

Tabla 10. *Resultados para los cambios de la base de datos de tipo intrónico profundo (tipo II) generados por los predictores (parte 3 de 3).*

Predictor **Cambio**	**Human Splicing Finder**	**NetGene2**	**NNSplice**	**Spliceman**	**SVM-BPfinder**	**Variant Effect Predictor tool**	**ESEfinder**
NTRK3 **c.2133+614A>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (69 %)	No efecto	Variante intrónica	Sitio *acceptor* ligeramente más débil
COL25A1 **c.1020+571C>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (77 %)	No efecto	Variante intrónica	No efecto
RTTN **c.5824-590G>A**	Alteración ratio ESE/ESS	No efecto	Sitio *acceptor* más fuerte	Alteración en el *splicing* (72 %)	No efecto	Variante intrónica	Sitio *acceptor* más débil
RBCK1 **c.1452+35C>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (47 %)	No efecto	Variante intrónica	No efecto
RAB11A **c.41-1373A>G**	Alteración ratio ESE/ESS. Activación *acceptor* críptico.	No efecto	No efecto	Alteración en el *splicing* (69 %)	No efecto	Variante intrónica	No efecto
TREML4 **c.507-1699A>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (85 %)	No efecto	Variante intrónica	Sitio *acceptor* más fuerte
ZMYND8 **c.3272-1970T>A**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (53 %)	No efecto	Variante intrónica	Sitio *donor* ligeramente más fuerte
BCAP29 **c.589+1714G>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (63 %)	No efecto	Variante intrónica	No efecto
VAV3 **c.321+1896C>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (94 %)	No efecto	Variante intrónica	No efecto

Tabla 11. *Resultados para los cambios de la base de datos de tipo exónico (tipo III o V) generados por los predictores (parte 1 de 3).*

Predictor **Cambio**	**Human Splicing Finder**	**NetGene2**	**NNSplice**	**Spliceman**	**SVM-BPfinder**	**Variant Effect Predictor tool**	**ESEfinder**	**EX-SKIP**	**HOT-SKIP**
PIK3CD **c.2808C>T**	Alteración ESE/ESS ratio	No efecto	No efecto	Alteración en el *splicing* (78%)	Aparición BP	Variante sinónima	Sitio *donor* y ESE debilitado	Posible exon *skipping*	No efecto
NCKAP1 **c.253C>T**	Alteración ESE/ESS ratio	No efecto	No efecto	Alteración en el *splicing* (66%)	No efecto	Variante sinónima	No efecto	Posible *skipping*	No efecto
CHEK2 **c.1116C>T**	No efecto	No efecto	No efecto	Alteración en el *splicing* (64%)	No efecto	Variante sinónima	No efecto	Posible *skipping*	No efecto
OR10H2 **c.612T>C**	No efecto	No efecto	No efecto	Alteración en el *splicing* (61%)	No efecto	Variante sinónima	No efecto	Posible *skipping*	Posible efecto (ratio 1)
TP53 **c.1024C>T**	No efecto	Nuevo sitio *acceptor*	No efecto	Alteración en el *splicing* (83%)	No efecto	Gana codón de *stop*	No efecto	Posible *skipping*	Posible efecto (ratio 3)
ELMO2 **c.1684C>T**	Activación sitio *acceptor* críptico	Nuevo sitio *acceptor*	No efecto	Alteración en el *splicing* (75%)	No efecto	Gana codón de *stop*	ESE más débiles	Posible *skipping*	No efecto
KRAS **c.38G>A**	Nuevo sitio *donor*	Nuevo sitio *acceptor*	No efecto	Alteración en el *splicing* (59%)	No efecto	Variante *missense*	Sitio *donor* y ESE más fuerte	No efecto	Posible efecto (ratio 0.25)
KCNJ11 **c.1054G>A**	No efecto	No efecto	No efecto	Alteración en el *splicing* (42 %)	No efecto	Variante *missense*	Pérdida sitio ESE	No efecto	No efecto
SORL1 **c.5914C>G**	Activación sitio *donor*	No efecto	No efecto	Alteración en el *splicing* (60%)	No efecto	Variante *missense*	Sitio *acceptor* debilitado	Posible *skipping*	Posible efecto (ratio 0.8)
BRCA1 **c.2106C>G**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (48%)	No efecto	Variante sinónima	Sitio *acceptor* y ESE debilitado	No se puede calcular	No efecto

Tabla 11. *Resultados para los cambios de la base de datos de tipo exónico (tipo III o V) generados por los predictores (parte 2 de 3).*

Predictor **Cambio**	**Human Splicing Finder**	**NetGene2**	**NNSplice**	**Spliceman**	**SVM-BPfinder**	**Variant Effect Predictor tool**	**ESEfinder**	**EX-SKIP**	**HOT-SKIP**
TP53 **c.92A>G**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (64 %)	No efecto	Variante *missense*	Sitio *acceptor* ligeramente más fuerte, cambios ESE	Posible *exon skipping*	No efecto
CLDN12 **c.*581T>C**	No se puede analizar	No efecto	No efecto	Alteración en el *splicing* (81 %)	No efecto	Variante *downstream* (3'UTR)	Sitio *donor* ligeramente más fuerte, cambios ESE	No se puede analizar	No efecto
KDR **c.802C>T**	Alteración ratio ESE/ESS	Sitio *acceptor* más débil	No efecto	Alteración en el *splicing* (53 %)	No efecto	Ganancia codón de parada	Sitio *acceptor* ligeramente más débil, cambios ESE	Posible *skipping*	No efecto
CHRNA7 **c.961G>T**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (61 %)	No efecto	Variante *missense*	Cambios ESE	Posible *skipping*	No efecto
KNG1 **c.1100A>G**	No efecto	No efecto	*Acceptor* más débil	Alteración en el *splicing* (70 %)	No efecto	Variante *missense*	No efecto	Posible *skipping*	No efecto
EPAS1 **c.2401C>A**	No efecto	No efecto	No efecto	Alteración en el *splicing* (71 %)	No efecto	Variante *missense*	Sitio *acceptor* más débil, cambios ESE	Posible *skipping*	No efecto
PGF **c.150C>T**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (50 %)	No efecto	Variante sinónima	Sitio *acceptor* ligeramente más débil, cambios ESE	Posible *skipping*	No efecto
NLRP1 **c.1308G>A**	Activación sitio *acceptor* críptico	Pérdida y aparición sitios *acceptor*	No efecto	Alteración en el *splicing* (73 %)	No efecto	Variante sinónima	Sitio *acceptor* activado, cambios ESE	No efecto	No efecto
NLRX1 **c.2627G>T**	Alteración ratio ESE/ESS	No efecto	No efecto	Alteración en el *splicing* (73 %)	No efecto	Variante *missense*	Sitio *acceptor* activado, cambios ESE	Posible *skipping*	No efecto
LMO7 **c.3281G>A**	No efecto	No efecto	No efecto	Alteración en el *splicing* (73 %)	No efecto	Variante missense	No efecto	No efecto	No efecto

Tabla 11. *Resultados para los cambios de la base de datos de tipo exónico (tipo III o V) generados por los predictores (parte 3 de 3).*

Predictor **Cambios**	**Human Splicing Finder**	**NetGene2**	**NNSplice**	**Spliceman**	**SVM-BPfinder**	**Variant Effect Predictor tool**	**ESEfinder**	**EX-SKIP**	**HOT-SKIP**
ADAM17 **c.359T>C**	Activación *donor* críptico	No efecto	No efecto	Alteración *splicing* (73%)	No efecto	Variante *missense* y de *splicing*	Sitio *donor* fortalecido y pérdida *donor*	No efecto	No efecto
TYW1 **c.270G>A**	No efecto	*Sitio donor* más débil	No efecto	Alteración *splicing* (72%)	No efecto	Variante sinónima	Sitio *acceptor* ligeramente más débil. Cambios ESE	No efecto	No efecto
JAK2 **c.1849G>T**	Alteración ESE/ESS	No efecto	No efecto	Alteración *splicing* (73%)	No efecto	Variante *missense*	Sitio *donor* más débil. Cambios ESE	Probab. *skipping*	Probabilidad *skipping* (19.0)
SMARCA4 **c.2843A>G**	Alteración ESE/ESS	No efecto	No efecto	Alteración *splicing* (65%)	No efecto	Variante *missense*	Pérdida *donor* y *acceptor* más fuerte. Cambios ESE.	No efecto	No efecto
KRAS **c.34G>C**	No efecto	Nuevos sitios *acceptor*	No efecto	Alteración *splicing* (81%)	No efecto	Variante *missense*	Pérdida sitio *donor.* Cambios ESE	No efecto	No efecto
KRAS **c.35G>A**	No efecto	Nuevos sitios *acceptor*	No efecto	Alteración *splicing* (64%)	No efecto	Variante *missense*	Pérdida sitio *donor.* Cambios ESE	No efecto	No efecto
CLIP4 **c.601G>A**	No efecto	Nuevo sitio *acceptor*	No efecto	Alteración *splicing* (54%)	No efecto	Variante *missense*	No efecto	No efecto	Probabilidad *skipping* (0.3)
TP53 **c.266G>A**	Activación *acceptor* críptico	No efecto	No efecto	Alteración *splicing* (78%)	No efecto	Variante *missense*	Cambios ESE	No efecto	No efecto
ARID1B **c.2262G>A**	Alteración ESE/ESS. Act. *donor* críptico	Sitio *acceptor* más fuerte	No efecto	Alteración *splicing* (73%)	BP nuevo	Variante sinónima	Cambios ESE	No efecto	Probabilidad *skipping* (4.00)

A continuación, se llevó el análisis de concordancia para cada uno de estos predictores para el total de las 99 variantes testeadas a lo largo de este trabajo. Primeramente, se estudió si el predictor producía algún tipo de efecto o no. También se estudió si la predicción coincidía o no con el tipo de efecto predicho por HSF. Los resultados se obtuvieron partiendo de la **Tabla 9**, **Tabla 10** y **Tabla 11**. El porcentaje de veces en que hay efecto para cada uno de los predictores, distinguiendo entre todas y por tipos de mutación, se muestra en la **Tabla 12**.

Tabla 12. *Porcentaje de veces en que hay efecto, distinguiendo entre todas las mutaciones y por cada tipo.*

	Porcentaje presencia efecto			
Predictor	**Todas las mutaciones**	**Mutaciones intrónicas cercanas**	**Mutaciones intrónicas profundas**	**Mutaciones exónicas**
Human Splicing Finder	71.70%	90.90%	63.60%	60.60%
NetGene2	43.40%	75.80%	24.20%	30.30%
NNSplice	32.30%	75.80%	18.20%	3.00%
Spliceman	73.70%	72.70%	72.70%	75.80%
SVM-BPfinder	15.20%	21.20%	15.20%	9.10%
Variant Effect Predictor tool	60.60%	100.00%	12.10%	69.70%
ESEfinder	67.70%	97.00%	33.30%	72.70%
EX-SKIP	-	-	-	51.70%
HOT-SKIP	-	-	-	24.10%

Seguidamente, se muestran los resultados de la tabla de confusión (**Tabla 13**), así como los estimadores para cada uno de los predictores para el estudio de la presencia o ausencia de efecto. Para el caso de EX-SKIP y HOT-SKIP, los valores corresponden solo para las mutaciones exónicas.

Tabla 13. *Resultados de la tabla de confusión cuando para el estudio de la presencia o no de efecto.*

Predictor	NetGene2	NNSplice	Spliceman	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder	EX-SKIP	HOT-SKIP
TP	35	29	53	14	48	57	10	4
TN	20	25	8	27	16	18	7	9
FP	8	3	20	1	12	10	5	3
FN	36	42	18	57	23	14	7	13
SENS.	49.30%	40.85%	74.65%	19.72%	67.61%	80.28%	58.82%	23.53%
SPEC.	71.43%	89.29%	28.57%	96.43%	57.14%	64.29%	58.33%	75.00%
PREC.	81.40%	90.63%	72.60%	93.33%	80.00%	85.07%	66.67%	57.14%
RECALL	49.30%	40.85%	74.65%	19.72%	67.61%	80.28%	58.82%	23.53%

TP: Verdaderos positivos; TN: Verdaderos negativos; FP: Falsos positivos; FN: Falsos ngativos; SENS.: Sensibilidad; SPEC: especificidad; PREC.: precision; RECALL: exhaustividad.

Las curvas ROC se generaron para cada uno de los predictores y se muestran a continuación:

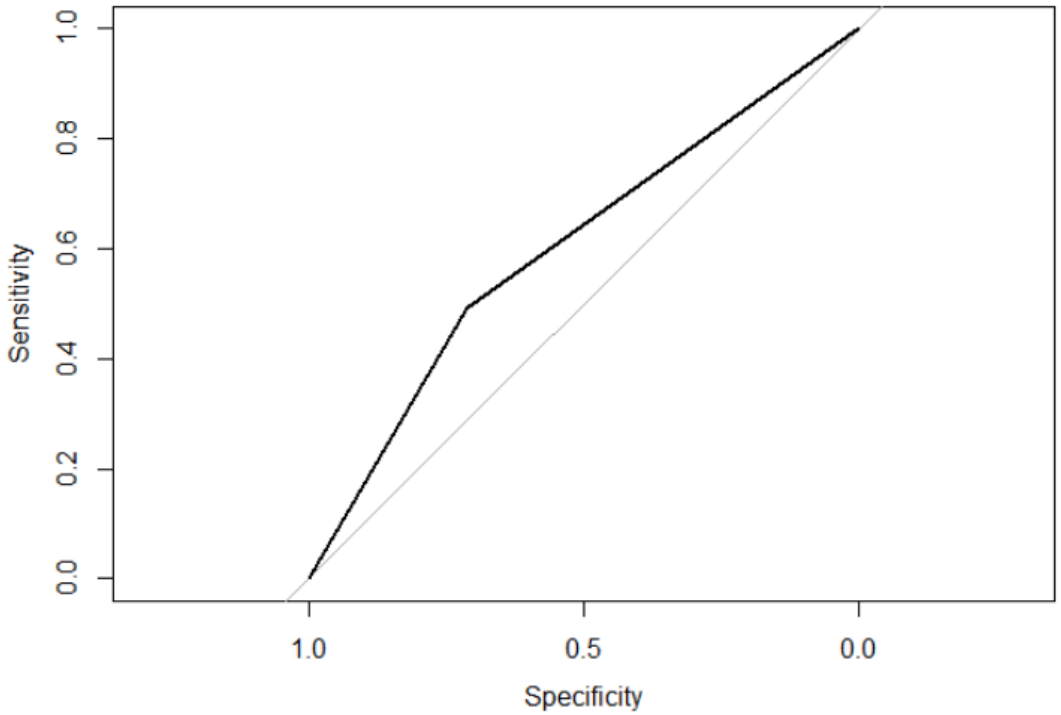

Figura 7. *Curva ROC para NetGene2 (Área bajo la curva: 0.6036).*

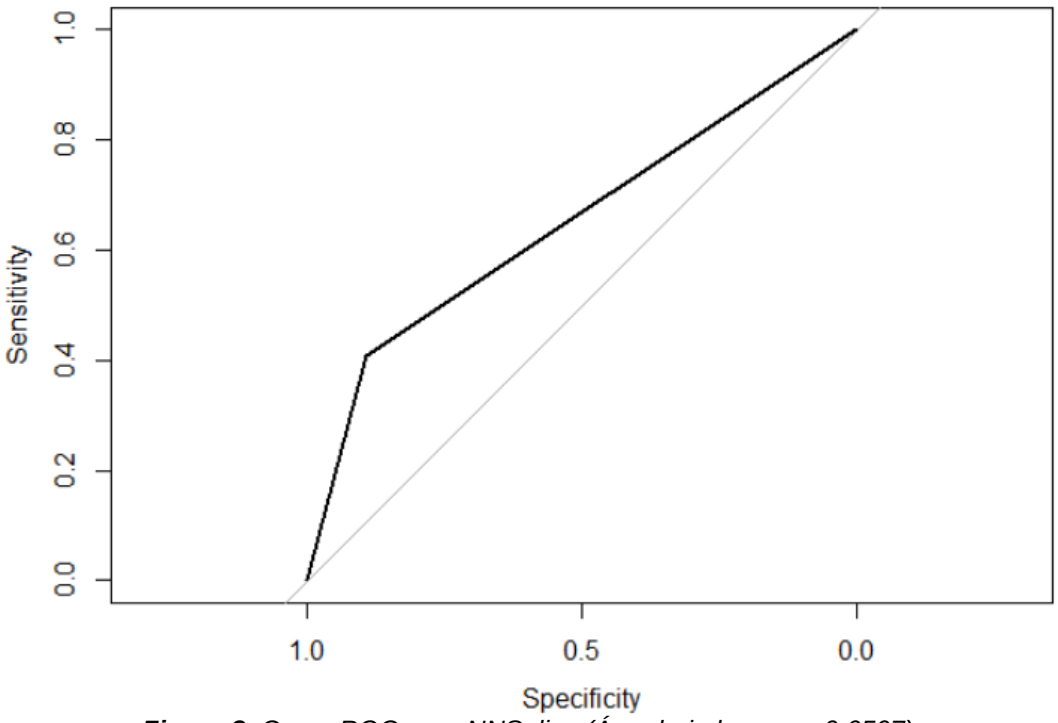

Figura 8. *Curva ROC para NNSplice (Área bajo la curva: 0.6507).*

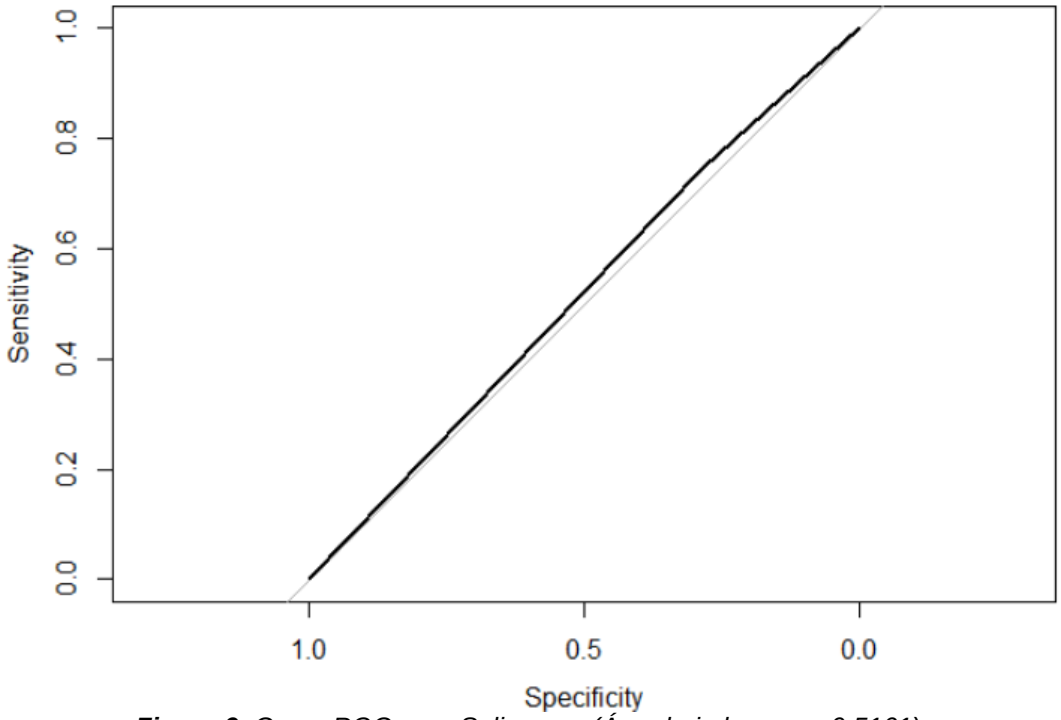

Figura 9. *Curva ROC para Spliceman (Área bajo la curva: 0.5161).*

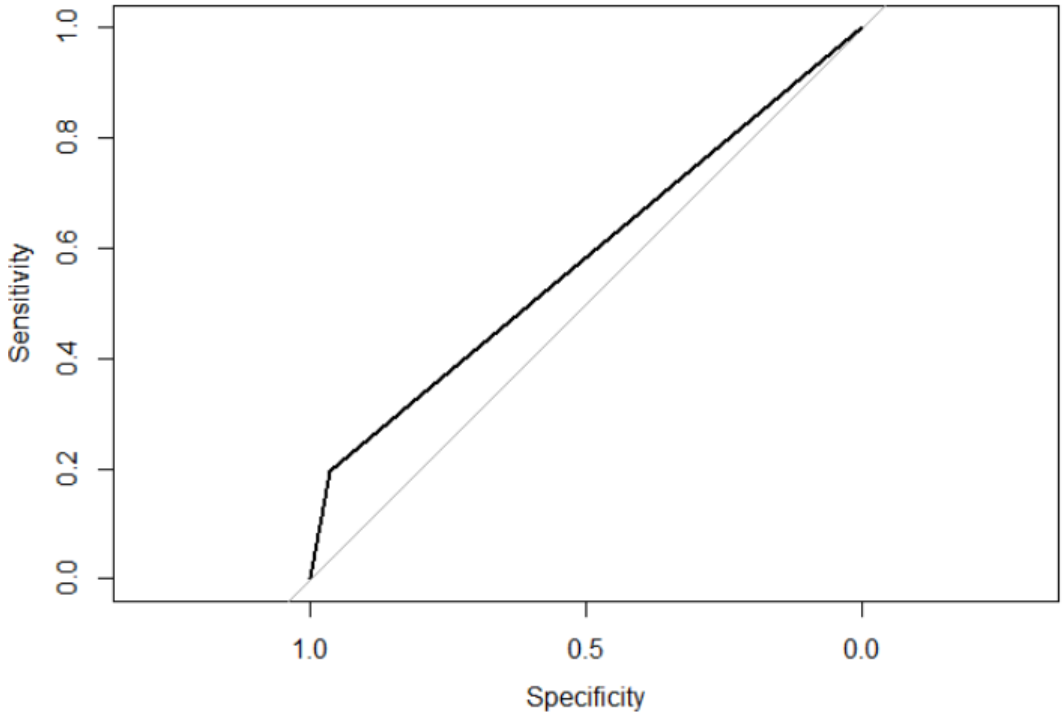

***Figura 10**. Curva ROC para SVM-BPfinder (Área bajo la curva: 0.5807).*

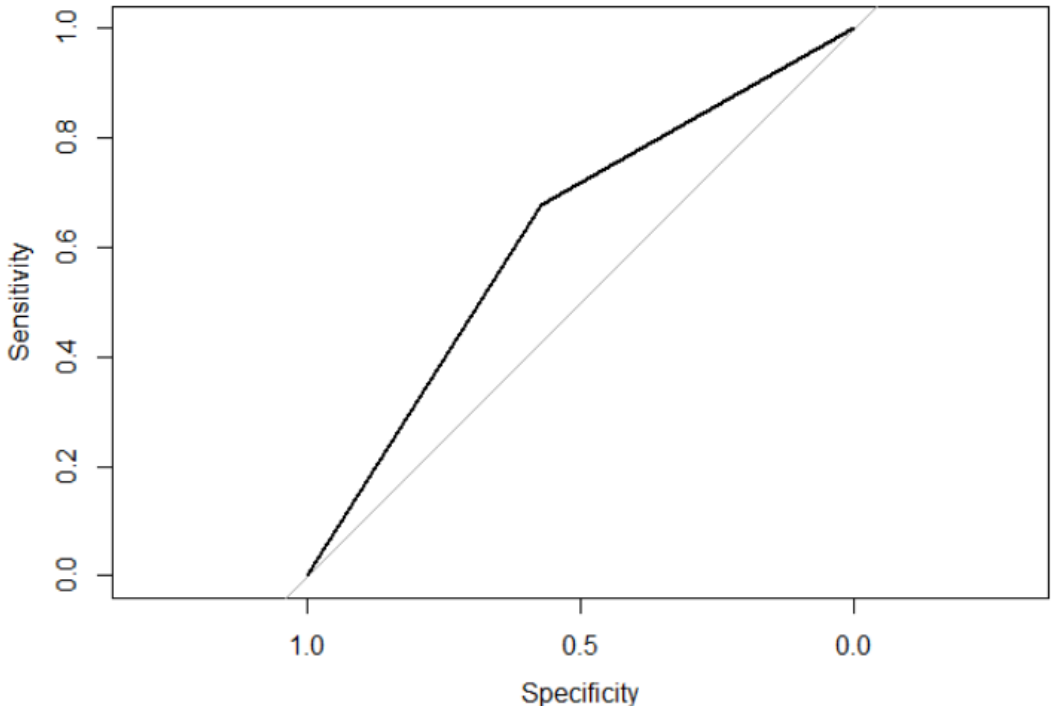

***Figura 11.** Curva ROC para Variant Effect Predictor tool (Área bajo la curva: 0.6237).*

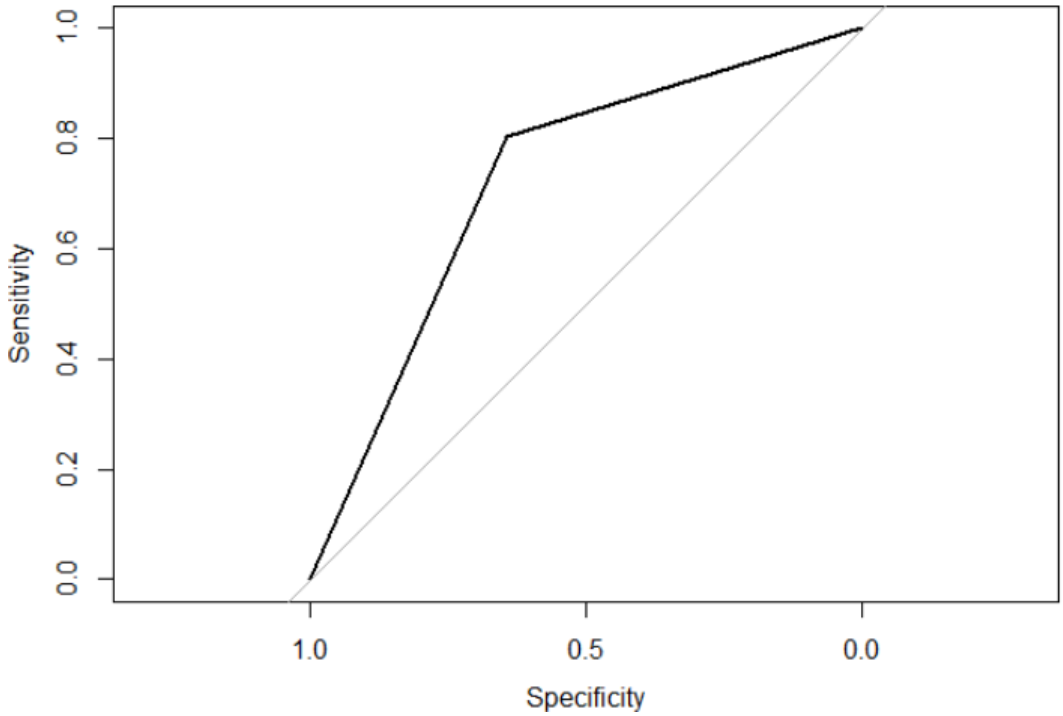

Figura 12. *Curva ROC para ESEfinder (Área bajo la curva: 0.7228).*

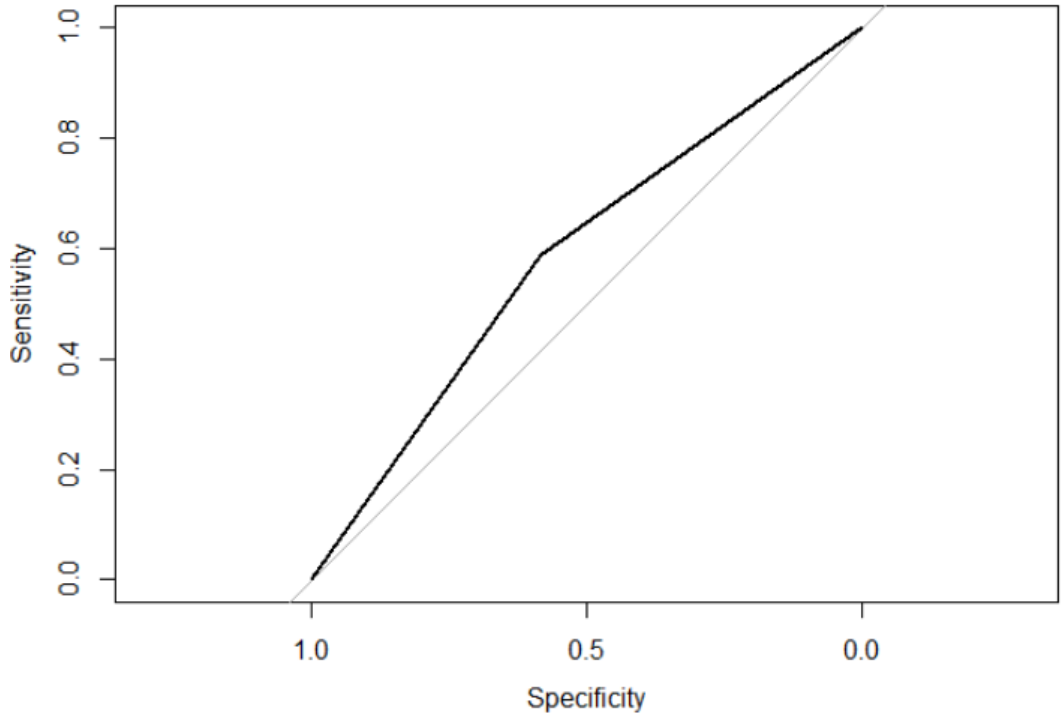

Figura 13. *Curva ROC para EX-SKIP (Área bajo la curva: 0.5858).*

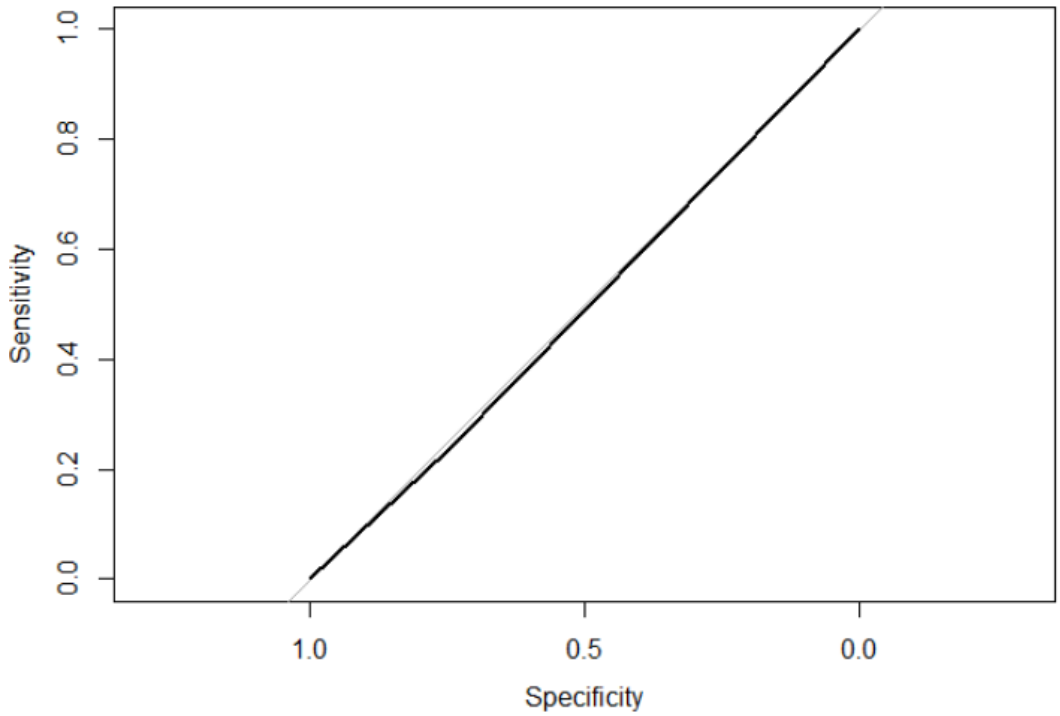

Figura 14. *Curva ROC para HOT-SKIP (Área bajo la curva: 0.4926).*

Seguidamente, se muestran los resultados para el análisis de concordancia (**Tabla 14**) y de la tabla de confusión para la comprobación de si la predicción coincide con el efecto predicho por HSF (**Tabla 15**). Para este paso, solo se tuvieron en cuenta los cambios que tenían algún tipo de efecto en este predictor (pasando de 99 cambios a 71), por lo que no se obtienen ni verdaderos negativos ni falsos positivos y, por tanto, tampoco se puede calcular la especificidad. Además, en este caso en que solo hay verdaderos positivos y falsos negativos, la sensibilidad y la exhaustividad serán igual a porcentaje de veces que hay efecto y la precisión es del 100% en todos los casos, ya que no hay falsos positivos. Para el caso de *EX-SKIP* y *HOT-SKIP*, los valores corresponden solo para las mutaciones exónicas. En el caso de *Spliceman*, como todos los resultados son negativos, no se puede hacer la tabla de confusión.

Tabla 14. *Porcentaje de veces que la predicción coincide con el efecto real para cada predictor, distinguiendo entre todas las mutaciones y por cada tipo.*

	Porcentaje acierto en efecto			
Predictor	**Todas las mutaciones**	**Mutaciones intrónicas cercanas**	**Mutaciones intrónicas profundas**	**Mutaciones exónicas**
NetGene2	36.60%	73.30%	9.50%	10.00%
NNSplice	35.20%	76.70%	9.50%	0.00%
Spliceman	0.00%	0.00%	0.00%	0.00%
SVM-BPfinder	11.30%	20.00%	4.80%	5.00%
Variant Effect Predictor tool	47.90%	100.00%	0.00%	20.00%
ESEfinder	64.80%	90.00%	19.00%	75.00%
EX-SKIP	-	-	-	11.80%
HOT-SKIP	-	-	-	17.60%

Tabla 15. *Resultados de la tabla de confusión para la comprobación si la predicción obtiene el efecto real o no.*

Predictor	NetGene2	NNSplice	SVM-BPfinder	Variant Effect Predictor tool	ESEfinder	EX-SKIP	HOT-SKIP
Verdaderos Positivos	26	25	8	34	46	2	3
Falsos Negativos	45	46	63	37	25	15	14
Sensibilidad	36.62%	35.21%	11.27%	47.89%	64.79%	11.76%	17.65%
Precisión	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Exhaustividad	36.62%	35.21%	11.27%	47.89%	64.79%	11.76%	17.65%

2.3 Discusión

Inicialmente, este trabajo tenía como objetivo conseguir un algoritmo que, a partir de una base de datos de mutaciones, se pudiera distinguir aquellas mutaciones que estuvieran relacionadas con el *splicing* y determinar qué efecto pueden tener sobre este fenómeno. Sin embargo, con la información contenida en la base de datos de partida no era posible conseguirlo, por lo que se debió cambiar el enfoque del trabajo. Partiendo de la misma base de datos, se seleccionarían una serie de variantes y se estudiaría la predicción que se obtiene para estas mutaciones en una serie de predictores *in silico*, seleccionados previamente y que habían sido analizados con cambios para los que se conocía qué efecto producen en el *splicing*.

Una de las partes centrales del trabajo fue la selección de los predictores *in silico* que se iban a emplear. Se buscaron diferentes referencias para encontrar herramientas en la red que predijeran el efecto de las mutaciones en el *splicing*. Muchas de las herramientas empleadas se desecharon en el momento en el que se observaron por las razones comentadas en los resultados (sección 2.2.2.2). Para el resto, con los cambios conocidos de la **tabla 2**, se observó qué predicciones generaban. Con los resultados de las **tablas 3** a **5**, se observa que no se pueden emplear *GENSCAN*, *MaxEntScan*, *CRYP-SKIP* e *IntSplice* para ninguno de los tipos de mutaciones, ya que no se están obteniendo resultados para ninguno de los tipos de cambios, por los que serán descartados y por ello no aparecen en los resultados de las **tablas 6** a **8**.

Con los resultados de estas últimas tablas, se observa que el predictor que mejor porcentaje de acierto tiene es *ESEFinder*, que no solo predijo efecto en las 12 variantes estudiadas (100.00%), sino que, aún más importante, predijo el efecto correcto en 8 de ellas (66.70%). El siguiente mejor es *Human Splicing Finder*, que solo falla en predecir efecto en una de las 12 variantes (91.70%) y acierta en el efecto en 8 (66.70%). Según estos resultados obtenidos para los

cambios conocidos, tanto *ESEfinder* como *Human Splicing Finder* acertaban a predecir el efecto real de la mutación en un 66.70% de los 12 casos estudiados. Además, se tenían referencias previas de la efectividad de *Human Splicing Finder* para la predicción del efecto de una variante en el *splcing* (Moles-Fernández *et al*. 2018), además de ofrecer el resultado del efecto de manera clara sin tener que interpretar los resultados. Por eso, se decidió emplear la herramienta HSF como referencia, o como equivalente al efecto real.

Un caso llamativo es el de *Variant Effect Predictor Tool*. El problema con esta herramienta es que no es clara a la hora de identificar el efecto de una mutación en el *splicing*. Donde más clara es, es en el caso de las mutaciones intrónicas cercanas, donde sí indica dónde puede estar afectando el cambio, aunque no diga en concreto el efecto que produce. Algo parecido ocurre con *Spliceman*, que indica que el cambio puede estar afectando al *splicing* en un porcentaje concreto, pero no indica qué efecto, por lo que no es demasiado útil para este estudio.

Estos resultados también muestran que *EX-SKIP* y *HOT-SKIP* no se pueden utilizar para las mutaciones intrónicas, ya que se utiliza para secuencias exónicas y, además, da resultados muy bajos. A pesar de que en una da resultado positivo, no quiere decir que sea por la mutación porque el cambio es muy pequeño.

Una vez reducido el número de predictores a emplear, se seleccionaron 29 variantes de cada uno de los tres archivos, obtenidas aleatoriamente a partir de la base de datos (con el *script* de *Python* del **Anexo II**). Esto se hizo así para tener el mismo número de mutaciones para los tres tipos (exónicas, intrónicas cercanas e intrónicas profundas) y no llegar al límite de pruebas que tenía el predictor *Human Splicing Finder* (100 búsquedas). Así, para el análisis del conjunto de variantes totales, se tendrán 33 variantes de cada tipo.

Las **tablas 9** a **11** reflejan los resultados de las predicciones para cada una de las variantes seleccionadas aleatoriamente de la base de datos. Al no conocer el efecto de las mutaciones estudiadas en el *splicing*, dado que no hay bibliografía al respecto, se debía tomar como referencia uno de los predictores que se habían empleado.

Con estos resultados se vuelve a comprobar los mismos resultados que se veían para los cambios conocidos. De los predictores, el que mejores resultados da es *ESEfinder*, igual que ocurría con los cambios conocidos, con un 64.80% de acierto en el efecto obtenido. *Spliceman* indica que el cambio puede estar relacionado con el *splicing*, pero no da el efecto y *Variant Effect Predictor Tool* tampoco indica efectos concretos, principalmente para las mutaciones intrónicas profundas. A pesar de no dar resultados positivos para las mutaciones exónicas, *NetGene2* da un resultado positivo, aunque no pasa lo mismo en *NNSplice*, que tenía un 0% de porcentaje para los cambios conocidos y este porcentaje no varía para todas las mutaciones de este tipo.

En cuanto a las curvas ROC, para el estudio de la presencia o no de efecto, son poco representativas, dado que no tienen la forma esperada ni valores muy elevados para el área bajo la curva. Por otro lado, estas no se pudieron realizar para el estudio de la coincidencia entre la predicción y el efecto de referencia porque no teníamos ejemplos de no efecto para HSF en esta parte del estudio.
Los resultados obtenidos en las predicciones pretendían extrapolarse y poder relacionar estos con los tipos de mutaciones explicadas por Wimmer *et al.* (2007). Sin embargo, esto no es posible llevarlo a cabo con todos los cambios estudiados debido a cómo expresan ciertas herramientas el efecto que pueden tener las variantes en el *splicing*, como se comentaba anteriormente con *Spliceman*. En los únicos predictores donde se puede establecer con cierta claridad qué tipo de mutaciones son es con *ESEfinder* y *Human Splicing Finder*. En el **Anexo VI** expone los resultados para las 99 variantes, según ambos predictores.

Según estos resultados, el 58.58% tienen predicciones que son iguales para ambas herramientas, siendo de estas el 36.21% mutaciones intrónicas cercanas (el 95.24% de tipo I y el 4.76% de tipo IV), el 18.97% mutaciones exónicas (el 36.36 de tipo III y el 63.64% de tipo V) y el 15.52% mutaciones intrónicas profundas (tipo II). Por lo tanto, de las variantes en las que coincide el efecto en ambos casos, el 70.69% están afectando al *splicing*. El resto (29.31%) son las variantes que no están causando ningún efecto en el *splicing*, según ambos predictores.

Con estos resultados, lo que se puede observar es que los predictores pueden identificar de una manera más precisa si las mutaciones intrónicas cercanas están afectando o no al *splicing*. Esto tiene cierta lógica porque estas se encuentran en las regiones correspondientes a los sitios *acceptor* y *donor* de los exones, por lo que cambios en estas regiones necesariamente tienen que estar afectando al *splicing*. El efecto más probable de estas mutaciones es que se rompa el sitio de *splicing* consenso, conduciendo a la pérdida del exón (tipo I) y, en menos ocasiones, que se produzca la inclusión de un fragmento de intrón o un salto de fragmento de exón (tipo IV).

Las mutaciones exónicas son más minoritarias. En mayor medida, afectan a las regiones reguladoras del *splicing,* tanto silenciadoras (ESS) como potenciadoras (ESE), que conducen a la omisión del exón completo (tipo V). Con menor frecuencia, estas mutaciones pueden activar sitios crípticos de *splicing*, ya sea *donor* o *acceptor*, conduciendo a la pérdida de un fragmento de exón (tipo III).

Las mutaciones intrónicas profundas son las más minoritarias, porque no solo es complicado que se produzca un cambio justo en regiones reguladoras que están muy alejadas del exón, así como generar nuevos sitios de *splicing* que, además, tengan cerca otro sitio que pueda generar la inclusión de un exón críptico en el *mRNA.*
De las mutaciones que no tienen efecto para ninguno de los dos predictores, el 64.71% son mutaciones intrónicas profundas y el 35.29% son mutaciones dentro de los exones. Esto nos indica lo que ya habíamos visto, que es difícil que las mutaciones en el interior de los intrones y alejadas de las regiones

codificantes puedan provocar cambios en el *splicing*, mientras que cualquier cambio en las intrónicas cercanas está generando una modificación del mensajero habitual.

En otros de los predictores, como ocurre con las herramientas *NetGene2* y *NNSplice*, puede haber confusión a la hora de determinar tipo de mutación, debido a que pueden ser dos tipos distintos. Por ejemplo, en el caso de las mutaciones intrónicas cercanas, la predicción puede indicar que se esté perdiendo un sitio *donor* o *acceptor*, lo que estaría indicando un cambio de tipo I, ya que se conduciría a la omisión del exón completo. Sin embargo, las herramientas, en muchas ocasiones, predicen más de un sitio *donor* o *acceptor* que hay en la secuencia que se ha dado al predictor. Estos sitios pueden estar empleándose, lo que daría un cambio de tipo IV, que conduciría a la inclusión de un fragmento de intrón o un salto de fragmento de exón. Por lo tanto, a pesar de que se sabe que hay un efecto en el *splicing*, no podemos saber claramente cuál de los dos tipos es.

En caso de agrupar las mutaciones en tipo I/IV, tipo II o tipo III/V, se obtienen los resultados que se muestran en el **Anexo VII**. En el 34.34% de las variantes, todos los predictores coinciden entre sí, siendo el 55.88% de tipo I/IV, el 8.82% de tipo II y el 35.29% no tienen efecto para ninguno de los predictores. En el 46.46%, *NetGene2* y *NNSplice* coinciden con la referencia (HSF), mientras que en el 39.39% del total de variantes coinciden ambos con *ESEfinder*. Estos resultados confirman lo que se había visto anteriormente. *NetGene2* y *NNSplice* sirven para predecir la pérdida o ganancia de sitios *donor* o *acceptor*, que ocurre mayoritariamente en las mutaciones intrónicas cercanas y en menor medida en las mutaciones intrónicas profundas. Sin embargo, la mayor parte de las mutaciones de tipo exónico afectan a regiones reguladoras, lo cual no es detectado por este tipo de predictores.

Todos estos resultados obtenidos, sin embargo, son predicciones. Como ya se había comentado anteriormente, las predicciones pueden ayudar a descartar el estudio cambios. Sin embargo, los resultados no son concluyentes hasta que no se analiza de manera experimental el efecto que pueden estar dando en el mRNA. Por lo tanto, el siguiente paso que se debe realizar una vez estudiado el efecto *in silico* es comprobar, bien con RT-PCR o con un ensayo de minigenes el efecto real de las mutaciones. Solo de esta manera se pueden mejorar los predictores actuales, ya que permite tener un amplio conjunto de variantes que servirían como datos para crear mejores algoritmos predictores en el futuro.

En resumen, los dos mejores predictores para indicar si las mutaciones están relacionadas con el *splicing* y los que más información dan sobre el efecto que va a tener dicha mutación sobre el mRNA son *Human Splicing Finder* y *ESEfinder*, que es el único que da resultados parecidos a la referencia de HSF. Estos deberían ser los predictores más a tener en cuenta a la hora de hacer un análisis de predicción de efecto de *splicing*.

3. Conclusiones

Se ha cumplido con el objetivo principal de este trabajo de estudiar los diferentes algoritmos bioinformáticos o predictores *in silico* para variantes genéticas que afectan al *splicing* dentro de un conjunto de variantes extraídas de un experimento de NGS y estudiar su nivel de precisión.

Partiendo de los objetivos secundarios planteados anteriormente y con los resultados obtenidos, las conclusiones que se obtienen en este TFM son las siguientes:

- Se ha hecho una revisión bibliográfica de métodos *in silico* de predicción de *splicing.*
- Se ha descargado una base de datos anotada de variantes genéticas relacionada con enfermedades humanas.
- Se ha interpretado la información de la anotación de variantes y extraído la información necesaria para averiguar si una variante afecta al *splicing,* a partir de la clasificación de Wimmer *et al.* (2007).
- Se han definido formalmente los criterios mediante los cuales una variante genética puede afectar al *splicing*.
- Se ha comparado entre predictores *in silico* de las variantes seleccionadas al azar.
- Se ha realizado un análisis descriptivo y estadístico de las predicciones obtenidas.
- El estudio de los predictores *in silico* nos permite identificar *Human Splicing Finder* y ESEfinder como las mejores herramientas para el diagnóstico de variantes relacionadas con el *splicing.*
- Muchos de los predictores *in silico* estudiados producen resultados que no pueden ser extrapolables a los tipos de *splicing* estudiados por Wimmer *et al.* (2007) o generan resultados que pueden llevar a concluir tipos distintos de mutaciones.
- Es imprescindible la validación experimental de los cambios para poder concluir de manera definitoria el verdadero efecto de estas mutaciones en el *splicing.*

Según los objetivos planteados, se ha realizado todo lo previsto en estos, a pesar que no se ha podido analizar ni tener todos los resultados deseados, debido a las complicaciones descritas anteriormente en referencia a la interpretación de los resultados de algunos de los predictores.

Según el plan de estudio fijado, se ha cumplido con los tiempos, incluso se ha conseguido terminar algunas de las partes con tiempo de antelación. La metodología podría haber estado mejor planteada desde un inicio, debido a que se tuvieron que hacer cambios en el enfoque del trabajo, pero, finalmente, se llegó a tener unos resultados concluyentes gracias a las modificaciones realizadas a tiempo.

Ya se ha comentado en la discusión que, con estos resultados, se debería analizar de manera experimental si verdaderamente son cambios relacionados

con el *splicing* y el efecto que tienen sobre el mRNA. Además, estos resultados también nos podrían ayudar en un futuro a crear un mejor *software* o predictor *in silico* para mejorar los que ya hay actualmente o que aporte algún punto de vista distinto no empleado hasta ahora.

En cuanto a las limitaciones del trabajo, se destacan las siguientes:

- Al partir únicamente de una base de datos donde no se indicaba el efecto en el *splicing*, se estaba probando a los predictores sin tener una referencia de lo que pasa realmente con esa mutación en el mRNA, no se puede probar la eficacia real de los predictores.
- Al tener que partir de uno de los predictores (HSF) como referencia para la comparación de los resultados, cuando se debería considerar como un candidato para ser el mejor predictor de los estudiados. Por lo tanto, las conclusiones se ven perjudicadas por ello. Además, el empleo de HSF limita el número de variantes que se pueden estudiar en este trabajo, debido a que solo admite testar 100 mutaciones.
- La interpretación de los resultados a partir de los predictores, ya que algunos no dicen el efecto que pueden estar teniendo en el *splicing*, mientras que otros pueden dar como resultado tipos diferentes de efecto, lo que no nos permite hacer clasificar los efectos de los predictores según Wimmer *et al.* (2007) para todo ellos.
- No poder testar los resultados obtenidos experimentalmente, lo que completaría el trabajo y permitiría confirmar el efecto de las mutaciones en *splicing*, haciendo que los datos sobre este fueran mayores y permitiera mejorar los predictores actuales.

En el caso de que, en un futuro, se quisiera retomar el trabajo planteado inicialmente, se necesitaría un *script* que fuera capaz de buscar en una base de datos que contuviera la secuencia del genoma completo la región donde se encuentra el cambio de interés y, a partir de una matriz de pesos donde se conociera cómo cada cambio afecta a las regiones de *splicing* o a las regiones reguladoras de este. Este *script* podría llevarse a cabo con *Python* o con algún otro programador más potente.

4. Glosario

Término	Definición
AUC Area Under the Curve (área bajo la curva)	Probabilidad de que un clasificador ordenará o puntuará una instancia positiva elegida aleatoriamente más alta que una negativa.
CDS Coding Sequence (Región codificante)	Porción del ADN de un gen o bien ARN que codifica la proteína. La región comienza en el extremo 5' por un codón de inicio y termina en el extremo 3' con un codón de stop.
DNA Deoxyribonucleic acid (ácido desoxirribonucleico)	Ácido nucleico que contiene las instrucciones genéticas usadas en el desarrollo y funcionamiento de todos los organismos vivos y algunos virus; también es responsable de la transmisión hereditaria.
ESE (exonic splicing enhancers)	Secuencia motivo de DNA de 6 bases dentro de un exón que dirige o mejora del mRNA.
ESS (exonic splicing silencers)	Región corta (generalmente 4-18 nucleótidos) de un exón que inhibe o silencia el splicing del mRNA y contribuyen al splicing alternativo.
Exón	Región de un gen que no es separada durante el proceso de corte y empalme y, por tanto, se mantienen en el ARN mensajero maduro. En los genes que codifican una proteína, son los exones los que contienen la información para producir la proteína codificada en el gen.
Intrón	Región del ADN que forma parte de la transcripción primaria de ARN, pero es eliminada del transcrito maduro previamente a su traducción.
ISE (intronic splicing enhancers)	Secuencia motivo de DNA de dentro de un intrón que dirige o mejora del mRNA.
ISS (intronic splicing silencers)	Región cortade un intrón que inhibe o silencia el splicing del mRNA y contribuyen al splicing alternativo.
mRNA Messenger RNA (ARN mensajero)	Ácido ribonucleico que transfiere el código genético procedente del DNA del núcleo celular a un ribosoma en el citoplasma y actúa como plantilla para la síntesis de una proteína.
Mutación	Cambio en una secuencia genética. Se emplea como sinónimo de variante.
NGS Next Generation Sequencing (secuenciación de nueva generación)	Tecnologías o técnicas diseñadas para analizar gran cantidad de ADN de forma masiva y paralela y que mejoran la secuenciación de Sanger.
NMD Nonsense-mediated decay (desintegración mediada sin sentido)	Mecanismo celular de vigilancia del ARN mensajero para detectar mutaciones terminadoras (PTC) y evitar la expresión de proteínas truncadas o erróneas.
ORF Open reading frame (marco de lectura abierto)	Secuencia de ARN comprendida entre un codón de inicio y un codón de stop, sin contar las secuencias que corresponden a los intrones.
PTC Premature termination codons (codones de stop	Codones de stop o parada que se producen como consecuencia de una mutación, lo que lleva a la traducción de proteínas más cortas.

prematuros)	
RNA Ribonucleic acid (ácido ribonucleico)	Ácido nucleico formado por una cadena de ribonucleótidos. Está presente tanto en las células procariotas como en las eucariotas, y es el único material genético de ciertos virus.
ROC Receiver Operating Characteristic (Característica Operativa del Receptor)	Representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación.
Spliceosome	Complejo molecular que se encuentra principalmente dentro del núcleo celular eucariota y que está formado por cinco ribonucleoproteínas nucleares pequeñas capaz de eliminar los intrones del ARNm durante el splicing.
Splicing	Proceso que ocurre entre la transcripción y la traducción que consiste en eliminar las regiones no codificantes del mRNA para producir un mRNA maduro.
SS Splicing Site (sitio de splicing)	Regiones de DNA canónicas y adyacentes a los exones, imprescindibles para el buen funcionamiento del splicing.

5. Bibliografía

Abramovicz, A., & Gos, M. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of applied genetics*, 59(3), 253–268.

Adzhubei, I., Jordan, D. M., & & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, Chapter 7, Unit7.20. .

Axelrod, F. B., Liebes, L., Gold-Von Simson, G., Mendoza, S., Mull, J., Leyne, M., y otros. (2011). Kinetin improves IKBKAP mRNA splicing in patients with familial dysautonomia. *Pediatric research*, 70(5), 480–483.

Baralle, D., & Baralle, M. (2005). Splicing in action: assessing disease causing sequence changes. *Journal of medical genetics* , 42(10), 737–748.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic acids research*, 28(1), 235–242.

Biamont, G., Catillo, M., Pignataro, D., Montecucco, A., & Ghigna, C. (2014). The alternative splicing side of cancer. *Seminars in cell & developmental biology*, 32, 30–36.

Brendel, V., Xing, L., & Zhu, W. (2004). Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. . *Bioinformatics (Oxford, England)*, 20(7), 1157–1169.

Brunak, S., Engelbrech, J., & Knudsen, S. (1991). Prediction of human mRNA donor and acceptor sites from the DNA sequence. . *J Mol Biol* , 220: 49–65.

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*, 268(1), 78–94.

Caminsky, N., Mucaki, E. J., & Rogan, P. K. (2014). Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *F1000Research*, 3, 282.

Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature reviews. Genetics*, 3(4), 285–298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., & Krainer, A. R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. . *Nucleic acids research*, 31(13), 3568–3571.

Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2011). Increased levels of noisy splicing in cancers, but not for oncogene-derived transcripts. . *Human molecular genetics*, 20(22), 4422–4429.

Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C. B., Ficarazzi, F., y otros. (2013). Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. . *PloS one*, 8(2), e57173.

Consortium, 1. G., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., y otros. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.

Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C., Green, E. D., Batzoglou, S., y otros. (2005). Distribution and intensity of constraint in mammalian genomic sequence. . *Genome research*, 15(7), 901–913.

Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and disease. . *Cell*, 136(4), 777–793. .

Corvelo, A., Hallegger, M., Smith, C. W., & Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. . *PLoS computational biology*, 6(11), e1001016.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563.

David, C. J., & Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & development,*, 24(21), 2343–2364.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., y otros. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. . *Human mutation*, 37(6), 564–569.

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. . *Nucleic acids research*, 37(9), e67.

Divina, P., Kvitkovicova, A., Buratti, E., & Vorechovsky, I. (2009). Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping. . *European journal of human genetics: EJHG*, 17(6), 759–765.

Dogan, R. I., Getoor, L., Wilbur, W. J., & Mount, S. M. (2007). SplicePort--an interactive splice-site analysis tool. . *Nucleic acids research*, 35(Web Server issue), W285–W291.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica. *Section D, Biological crystallography*, 60(Pt 12 Pt 1), 2126–2132.

Eng, L., Coutinho, G., Nahas, S., Yeo, G., Tanouye, R., Babaei, M., y otros. (2004). Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Human mutation*, 23(1), 67–76.

Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J. O., & Schaal, H. (2014). Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. . *Nucleic acids research*, 42(16), 10681–10697.

Fairbrother, W. G., Yeh, R. F., Sharp, P. A., & Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. . *Science (New York, N.Y.)*, 297(5583), 1007–1013.

Fang, L. J., Simard, M. J., Vidaud, D., Assouline, B., Lemieux, B., Vidaud, M., y otros. (2001). A novel mutation in the neurofibromatosis type 1 (NF1) gene promotes skipping of two exons by preventing exon definition. . *Journal of molecular biology*, 307(5), 1261–1270.

Faustino, N. A., & Cooper, T. A. (2003). Pre-mRNA splicing and human disease. . *Genes & development*, 17(4), 419–437. .

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., y otros. (2017). COSMIC: somatic cancer genetics at high-resolution. . *Nucleic acids research*, 45(D1), D777–D783.

Fredericks, A. M., Cygan, K. J., Brown, B. A., & Fairbrother, W. G. (2015). RNA-Binding Proteins: Splicing Factors and Disease. . *Biomolecules*, 5(2), 893–909. .

Gilad, S., Khosravi, R., Shkedy, D., Uziel, T., Ziv, Y., Savitsky, K., et al. (1996). Predominance of null mutations in ataxia-telangiectasia. *Human molecular genetics*, 5(4), 433–439.

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582(14), 1977–1986.

Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., y otros. (2006). Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. . *Molecular cell*, 22(6), 769–781.

Gulley, M. L., Braziel, R. M., Halling, K. C., Hsi, E. D., Kant, J. A., Nikiforova, M. N., y otros. (2007). Clinical laboratory reports in molecular pathology. *Archives of pathology & laboratory medicine*, 131(6), 852–863.

Hebsgaard, S. M., Korning, P. G., Tolstrup, N., Engelbrecht, J., Rouze, P., & Brunak, S. (1996). Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. *Nucleic Acids Res* , 24(17):3439–3452.

Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve? *Emergency medicine journal : EMJ*, 34(6), 357–359. .

Hori, T., Fukao, T., Murase, K., Sakaguchi, N., Harding, C. O., & Kondo, N. (2013). Molecular basis of two-exon skipping (exons 12 and 13) by c.1248+5g>a in OXCT1 gene: study on intermediates of OXCT1 transcripts in fibroblasts. . *Human mutation*, 34(3), 473–480.

Ibrahim, E. C., Hims, M., Shomron, N., Burge, C. B., Slaugenhaupt, S. A., & Reed, R. (2007). Weak definition of IKBKAP exon 20 leads to aberrant splicing in familial dysautonomia. *Hum Mutat*, 28:41–5.

Känsäkoski, J., Jääskeläinen, J., Jääskeläinen, T., Tommiska, J., Saarinen, L., Lehtonen, R., y otros. (2016). Complete androgen insensitivity syndrome caused by a deep intronic pseudoexon-activating mutation in the androgen receptor gene. *Scientific reports*, 6, 32819. .

Kim, E., Goren, A., & Ast, G. (2008). Insights into the connection between cancer and alternative splicing. *Trends in genetics: TIG*, 24(1), 7–10.

Koch, L. (2020). Exploring human genomic diversity with gnomAD. *Nature reviews. Genetics*, 21(8), 448.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., y otros. (2016). Analysis of protein-coding genetic variation in 60,706 humans. . *Nature*, 536(7616), 285–291.

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 189–1.

Lim, K. H., & Fairbrother, W. G. (2012). Spliceman--a computational web server that predicts sequence variations in pre-mRNA splicing. *Bioinformatics (Oxford, England)*, 28(7), 1031–1032.

Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., & Fairbrother, W. G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. . *Proceedings of the National Academy of Sciences of the United States of America*, 108(27), 11093–11098. .

McAlinden, A., Majava, M., Bishop, P. N., Perveen, R., Black, G. C., Pierpont, M. E., y otros. (2008). Missense and nonsense mutations in the alternatively-spliced exon 2 of COL2A1 cause the ocular variant of Stickler syndrome. . *Human mutation*, 29(1), 83–90.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. . *Bioinformatics (Oxford, England)*, 26(16), 2069–2070.

Modrek, B., & Lee, C. (2002). A genomic view of alternative splicing. *Nature genetics*, 30(1), 13–19.

Moles-Fernández, A., Duran-Lozano, L., Montalban, G., Bonache, S., López-Perolio, I., Menéndez, M., y otros. (2018). Computational Tools for Splicing Defect Prediction in Breast/Ovarian Cancer Genes: How Efficient Are They at Predicting RNA Alterations? *Frontiers in genetics*, 9, 366.

Nature. (2018). *Mutation | Learn Science at Scitable*. Obtenido de https://www.nature.com/subjects/mutation

Ng, P. C., & & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. . *Nucleic acids research*, 31(13), 3812–3814.

Nissim-Rafinia, M., & Kerem, B. (2002). Splicing regulation as a potential genetic modifier. *Trends in genetics: TIG*, 18(3), 123–127.

Palhais, B., Dembic, M., Sabaratnam, R., Nielsen, K. S., Doktor, T. K., Bruun, G. H., et al. (2016). The prevalent deep intronic c. 639+919 G>A GLA mutation causes pseudoexon activation and Fabry disease by abolishing the binding of hnRNPA1 and hnRNP A2/B1 to a splicing silencer. . *Molecular genetics and metabolism*, 119(3), 258–269.

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. . *Nucleic acids research*, 29(5), 1185–1190.

Piva, F., Giulietti, M., Burini, A. B., & Principato, G. (2012). SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Human mutation*, 33(1), 81–85.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. . *Genome research*, 20(1), 110–121. .

Rahman, M., Nasrin, F., Masuda, A., & Ohno, K. (2015). Decoding Abnormal Splicing Code in Human Diseases. *Journal of Investigative Genomics*, 2(1), 6-23.

Ramalho, A. S., Beck, S., Penque, D., Gonska, T., Seydewitz, H. H., Mall, M., y otros. (2003). Transcript analysis of the cystic fibrosis splicing mutation 1525-1G>A shows use of multiple alternative splicing sites and suggests a putative role of exonic splicing enhancers. *Journal of medical genetics*, 40(7), e88.

Raponi, M., Kralovicova, J., Copson, E., Divina, P., Eccles, D., Johnson, P., y otros. (2011). Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Human mutation*, 32(4), 436–444.

Raraigh, K. S., Han, S. T., Davis, E., Evans, T. A., Pellicore, M. J., McCague, A. F., y otros. (2018). Functional Assays Are Essential for Interpretation of

Missense Variants Associated with Variable Expressivity. *American journal of human genetics*, 102(6),1062-1077.

Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. . *Journal of computational biology : a journal of computational molecular cell biology, Journal of computational biology: a journal of computational molecular cell biology*, 4(3), 311–323. .

Rhine, C. L., Cygan, K. J., Soemedi, R., Maguire, S., Murray, M. F., Monaghan, S. F., et al. (2018). Hereditary cancer genes are highly susceptible to splicing mutations. *PLoS genetics*, 14(3), e1007231. .

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., y otros. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. . *Genetics in medicine : official journal of the American College of Medical Genetics*, 17(5), 405–424.

Rogozin, I. B., & & Milanesi, L. (1997). Analysis of donor splice sites in different eukaryotic organisms. . *Journal of molecular evolution*, 45(1), 50–59. .

Sanz, D. J., Hollywood, J. A., Scallan, M. F., & Harrison, P. T. (2017). Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA. *PloS one*, 12(9), e0184009.

Schwartz, S., Hall, E., & Ast, G. (2009). SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. . *Nucleic acids research*, 37(Web Server issue), W189–W192.

Shapiro, M. B., & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic acids research*, 15(17), 7155–7174.

Sharma, N., Sosnay, P. R., Ramalho, A. S., Douville, C., Franca, A., Gottschalk, L. B., y otros. (2014). Experimental assessment of splicing variants using expression minigenes and comparison with in silico predictions. *Human mutation*, 35(10), 1249–1259.

Shibata, A., Okuno, T., Rahman, M. A., Azuma, Y., Takeda, J., Masuda, A., y otros. (2016). IntSplice: prediction of the splicing consequences of intronic single-nucleotide variations in the human genome. . *Journal of human genetics*, 61(7), 633–640.

Singh, G., & Cooper, T. A. (2006). Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing. . *BioTechniques*, 41(2), 177–181.

Smith, P. J., Zhang, C., Wang, J., Chew, S. L., Zhang, M. Q., & Krainer, A. R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. . *Human molecular genetics*, 15(16), 2490–2508.

Sobczyńska-Tomaszewska, A., Ołtarzewski, M., Czerska, K., Wertheim-Tysarowska, K., Sands, D., Walkowiak, J., y otros. (2013). Newborn screening for cystic fibrosis: Polish 4 years' experience with CFTR sequencing strategy. *European journal of human genetics: EJHG*, 21(4), 391–396.

Stenson, P. D., Ball, E. V., Chapman, M., Evans, K., Azevedo, L., Hayden, M., et al. (2020). The Human Gene Mutation Database (HGMD®): optimizing

its use in a clinical diagnostic or research setting. *Human genetics*, 139(10),1197-1207.

Sterne-Weiler, T., & Sanford, J. R. (2014). Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome biology*, 15(1), 201.

Svaasand, E. K., Engebretsen, L. F., Ludvigsen, T., Brechan, W., & Sjursen, W. (2015). A Novel Deep Intronic Mutation Introducing a Cryptic Exon Causing Neurofibromatosis Type 1 in a Family with Highly Variable Phenotypes: A Case Study. . *Hereditary Genet* , 4: 152. .

Symoens, S., Malfait, F., Vlummens, P., Hermanns-Lê, T., Syx, D., & De Paepe, A. (2011). A novel splice variant in the N-propeptide of COL5A1 causes an EDS phenotype with severe kyphoscoliosis and eye involvement . *PloS one*, 6(5), e20121. .

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., y otros. (2019). COSMIC: the Catalogue og Somatic Mutations in Cancer. *Nucleic acids research*, 47(D1), D941-D947.

Tazi, J., Bakkour, N., & Stamm, S. (2009). Alternative splicing and disease. *Biochimica et biophysica acta*, 1792(1), 14–26.

Urbanski, L. M., Leclair, N., & Anczuków, O. (2018). Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. Wiley interdisciplinary reviews. *RNA*, 9(4), e1476.

Vaz-Drago, R., Custódio, N., & Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Human genetics*, 136(9), 1093–1111.

Vázquez, C. (2019). Bases genéticas de las ataxias. Análisis de mutaciones de splicing en ATM (Trabajo de Final de Grado). *Universitat Politècnica de València, Valencia, España*.

Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. . *Nature reviews. Genetics*, 8(10), 749–761.

Wang, Z., Jensen, M. A., & Zenklusen, J. C. (2016). A practical guide to the cancer genome atlas (TCGA). . *In Statistical Genomics* , pp. 111-141.

Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., & Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. . *Cell*, 119(6), 831–845.

Ward, A. J., & Cooper, T. A. (2010). The pathobiology of splicing. . *The Journal of pathology* , 220(2), 152–163.

Wimmer, K., Roca, X., Beiglböck, H., Callens, T., Etzler, J., Rao, A. R., et al. (2007). Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5' splice-site disruption. . *Human mutation*, 28(6), 599–612.

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., y otros. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. . *Science (New York, N.Y.)*, 347(6218), 1254806.

Xu, W., Yang, X., Hu, X., & Li, S. (2014). Fifty-four novel mutations in the NF1 gene and integrated analyses of the mutations that modulate splicing. *International journal of molecular medicine*, 34(1), 53–60.

Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. . *Journal of*

computational biology: a journal of computational molecular cell biology, 11(2-3), 377–394.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., et al. (2018). Ensembl 2018. *Nucleic acids research*, 46(D1),D754-D761.

Zhang, X. H., & Chasin, L. A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. . *Genes & development*, 18(11), 1241–1250.

6. Anexos

ANEXO I. *Script de Python* que separa las mutaciones entre exónicas, Intrónicas cercanas e intrónicas profundas.

```
import sys

# importamos el fichero con el que vamos a trabajar
file = sys.argv[1]

# importamos el nombre de la carpeta de salida por pantalla
folder = sys.argv[2]

# abrir fichero para lectura
infile = open(file)

# ficheros de salida para los diferentes tipos de mutaciones
folder = folder.replace('\\','/')

file1 = folder + '/Mutations_I.tsv'
file2 = folder + '/Mutations_II.tsv'
file3 = folder + '/Mutations_III.tsv'
file4 = folder + '/Mutations_extra.tsv'

outfile1 = open(file1,"w") # tipo I/IV
outfile2 = open(file2,"w") # tipo II
outfile3 = open(file3,"w") # tipo III/V
outfile4 = open(file4,"w") # mutaciones que tienen nomenclatura extraña

# escribir nombres de las columnas
outfile1.write('Gene name'+'\t'+'Mutation CDS'+'\t'+'Mutation Description'+'\n')
outfile2.write('Gene name'+'\t'+'Mutation CDS'+'\t'+'Mutation Description'+'\n')
outfile3.write('Gene name'+'\t'+'Mutation CDS'+'\t'+'Mutation Description'+'\n')
outfile4.write('Gene name'+'\t'+'Mutation CDS'+'\t'+'Mutation Description'+'\n')

def RepresentsInt(s):
   try:
       int(s)
       return True
   except ValueError:
       return False

i = 1
# para cada linea del fichero
for line in infile:
    # quitamos el salto de línea y separamos las columnas delimitadas por tabuladores
    columns = line.strip('\n').split('\t')
    # tomamos el primer campo, que es el nombre del gen
    genname = columns[0]
    # tomamos el campo 20, que es la anotación
    anotation = columns[19]
    # tomamos el campo 22, que es la descripción
    description = columns[21]

    # si es un cambio de nucleótidos, tendrá '>'
    if '>' in anotation:
        # si empieza por 'c.-', no nos interesa
        if ('c.-' in anotation):
            # dejarla pasar
            continue
        # si no empieza por 'c.-'
        else:
            # si tiene '+' será mutación intrónica downstream
            if ('+' in anotation):
```

```
            # separamos por el '>' y tomamos la parte de la izquierda
            # separamos por el '+' y tomamos la parte de la derecha
            # nos desacemos de la letra caracter para quedarnos con el número
            number = anotation.split('>')[0].split('+')[1][:-1]

            if not RepresentsInt(number):
               print ("Warning: " + number + 'in line' + str(i) + " not an integer.")
               outfile4.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')
               continue

            # si el número es mayor a 6
            if (int(number) > 6):
                # será una variable deep intronic (tipo II)
                outfile2.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')
            # si no
            else:
                # será una variable intrónica cercana al exón (tipo I o IV)
                outfile1.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')

        # si tiene tiene '-' será mutacion intrónica upstream
        elif ('-' in anotation):
            # separamos por el '>' y tomamos la parte de la izquierda
            # separamos por el '-' y tomamos la parte de la derecha
            # nos desacemos de la letra para quedarnos con el número
            number = anotation.split('>')[0].split('-')[1][:-1]

            if not RepresentsInt(number):
               print ("Warning: " + number + 'in line' + str(i) + " not an integer.")
               outfile4.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')
               continue

            # si el número es mayor a 6
            if (int(number) > 6):
                # será una variable deep intronic (tipo II)
                outfile2.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')
            # si no
            else:
                # será una variable intrónica cercana al exón (tipo I o IV)
                outfile1.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')

        # si no tiene ni '+' ni '-', será una mutación exónica
        else:
            # será una mutación de tipo III o IV
            outfile3.write(columns[0]+'\t'+columns[19]+'\t'+columns[21]+'\n')
# si no es un cambio nucleotídico
else:
    # dejar pasar la linea
    continue
print(i)
i+=1
```

ANEXO II. *Script* de *Python* que selecciona muestras de manera aleatoria de los ficheros de mutaciones.

```
import random

random_lines = random.choices(open('Mutations_I.tsv').readlines(),k=29)


outfile=open('MutationsI_sample.tsv','w')
outfile.write('Gene name'+'\t'+'Transcript ID'+'\t'+'Mutation CDS'+'\t'+'Chromosomic Position'+'\n')

for i in random_lines:
    outfile.write(i)
    print(i)


outfile.close()
```

Figura 15. *Script de Python que selecciona de manera aleatoria 29 mutaciones del fichero para las mutaciones intrónicas cercanas.*

```
import random

random_lines = random.choices(open('Mutations_II.tsv').readlines(),k=29)


outfile=open('MutationsII_sample.tsv','w')
outfile.write('Gene name'+'\t'+'Transcript ID'+'\t'+'Mutation CDS'+'\t'+'Chromosomic Position'+'\n')

for i in random_lines:
    outfile.write(i)
    print(i)


outfile.close()
```

Figura 16. *Script de Python que selecciona de manera aleatoria 29 mutaciones del fichero para las mutaciones intrónicas profundas.*

```
import random

random_lines = random.choices(open('Mutations_III.tsv').readlines(),k=29)


outfile=open('MutationsIII_sample.tsv','w')
outfile.write('Gene name'+'\t'+'Transcript ID'+'\t'+'Mutation CDS'+'\t'+'Chromosomic Position'+'\n')

for i in random_lines:
    outfile.write(i)
    print(i)


outfile.close()
```

Figura 17. *Script de Python que selecciona de manera aleatoria 29 mutaciones del fichero para las mutaciones exónicas.*

ANEXO III. Resultados completos para los cambios analizados.

El enlace siguiente conduce a una carpeta con los resultados para cada uno de los 99 cambios analizados en cada uno de los predictores: https://drive.google.com/drive/folders/1ZHfKhHSQmeCwyWQFffRLY9NA4c4qtn1Z?usp=sharing

ANEXO IV. Resultados para la comparación entre las predicciones obtenidas y el efecto real de los cambios conocidos (**Tabla 2**).

	cambio	tipo	efecto	netgene	nnsplice	genscan	maxenstsc	splicema	cryp	hsf	svm	intsp	vept	esefinc	exskip	hotskip
2	ATM c.2921+1G>A	Tipo I/IV	1	0	1	0	0	1	0	1	1	0	1	1	0	0
3	NF1 c.1845+1G>A	Tipo I/IV	1	0	1	0	0	1	0	1	0	0	1	1	1	0
4	COL5A1 c.925-2A>G	Tipo I/IV	1	0	1	0	0	1	0	1	0	0	1	1	0	0
5	OXCT1 c.1248+5G>A	Tipo I/IV	1	1	1	0	0	1	0	1	1	0	1	1	0	0
6	NF1 c.288+1137G>A	Tipo II	1	1	1	0	0	1	0	1	0	0	1	1	0	0
7	CFTR c.3718-2477C>T	Tipo II	1	1	1	0	0	1	0	1	0	0	1	1	0	0
8	AR c.2450-118A>G	Tipo II	1	1	1	0	0	1	0	1	0	0	1	1	0	0
9	GLA c.639+919G>A	Tipo II	1	1	0	0	0	1	0	1	1	0	1	1	0	0
10	COL2A1 c.192G>A	Tipo III/V	1	0	0	0	0	1	0	1	0	0	1	1	1	0
11	ACADM c.382C>T	Tipo III/V	1	0	0	0	0	0	0	0	0	0	1	1	1	0
12	NF1 c.3362A>G	Tipo III/V	1	0	0	0	0	0	0	1	0	0	1	1	1	1
13	BRCA1 c.4484G>T	Tipo III/V	1	0	0	0	0	1	0	1	1	0	1	1	0	0

Figura 18. *Comparación entre predictores y efecto real. Se indica con un 1 si hay efecto y con un 0 si no hay efecto.*

	cambio	tipo	efecto	netgene2	nnsplice2	genscan2	maxenstsc	splicema	cryp	hsf2	svm2	intspl	vept2	esefind	exskip2	hotskip2
2	ATM c.2921+1G>A	Tipo I/IV	1	0	1	0	0	0	0	1	1	0	1	1	0	0
3	NF1 c.1845+1G>A	Tipo I/IV	1	0	1	0	0	0	0	1	0	0	1	1	1	0
4	COL5A1 c.925-2A>G	Tipo I/IV	1	0	1	0	0	0	0	1	0	0	1	1	0	0
5	OXCT1 c.1248+5G>A	Tipo I/IV	1	1	1	0	0	0	0	1	1	0	1	1	0	0
6	NF1 c.288+1137G>A	Tipo II	1	1	1	0	0	0	0	1	0	0	0	1	0	0
7	CFTR c.3718-2477C>T	Tipo II	1	1	1	0	0	0	0	0	0	0	0	1	0	0
8	AR c.2450-118A>G	Tipo II	1	1	1	0	0	0	0	1	0	0	0	1	0	0
9	GLA c.639+919G>A	Tipo II	1	1	0	0	0	0	0	1	1	0	0	0	0	0
10	COL2A1 c.192G>A	Tipo III/V	1	0	0	0	0	0	0	0	0	0	0	0	1	0
11	ACADM c.382C>T	Tipo III/V	1	0	0	0	0	0	0	0	0	0	0	1	1	0
12	NF1 c.3362A>G	Tipo III/V	1	0	0	0	0	0	0	1	0	0	1	0	1	1
13	BRCA1 c.4484G>T	Tipo III/V	1	0	0	0	0	0	0	0	1	0	0	0	0	0

Figura 19. *Comparación entre predictores y efecto real. Se indica con un 1 si la predicción coincide con el efecto real y con un 0 si no coinciden.*

ANEXO V. Variantes seleccionadas al azar de los ficheros obtenidos de ejecutar el *script* del **Anexo I**.

Tabla 16. *Variantes seleccionadas de los ficheros generados por el script del* ***Anexo I*** *(Parte 1 de 2).*

Gen	Transcrito	Anotación	Posición cromosómica	Gen	Transcrito	Anotación	Posición cromosómica
CACNA1C	ENST00000402845.7	c.1218-2A>G	12:2512810	**DIDO1**	ENST00000395340.5	c.2214+2T>C	20:62896231
GIT1	ENST00000581348.5	c.405+5G>C	17:29582693	**ARHGAP28**	ENST00000383472.8	c.955-1G>T	18:6873408
CHD4	ENST00000642879.1	c.3820-1G>T	12:6583379	**FRMPD4**	ENST00000380682.5	c.813+3C>G	23:12690329
PIR	ENST00000380420.9	c.273+1G>A	23:15459656	**NIN**	ENST00000453196.5	c.1546-2A>G	14:50766398
UEVLD	ENST00000320750.10	c.821-1G>C	11:18544797	**CLASRP**	ENST00000544944.6	c.100-1G>A	19:45052070
CTCFL	ENST00000432255.6	c.755-1G>T	20:57519378	**CHL1**	ENST00000620033.4	c.197+1G>A	3:326065
NF2	ENST00000403435.5	c.1488-1G>C	22:29681438	**NPR3**	ENST00000415167.2	c.892+1G>T	5:32724821
U2AF1	ENST00000291552.8	c.44+1G>A	21:43107450	**MFSD8**	ENST00000641743.1	c.554-1G>T	4:127939998
MAP4K4	ENST00000324219.8	c.640-1G>T	2:101834408	**SEMA4A**	ENST00000368285.7	c.685+1G>A	1:156160560
RHCE	ENST00000413854.5	c.939+1G>T	1:25388975	**GFPT2**	ENST00000253778.12	c.2004+4C>T	5:180302419
TP53	ENST00000445888.6	c.375+1G>A	17:7675993	**PTPRT**	ENST00000373193.7	c.1865+1G>T	20:42350627
KLHL2	ENST00000514860.5	c.1480+2T>C	4:165313368	**STK11**	ENST00000586243.5	c.465-1G>T	19:1220372
TSPAN17	ENST00000508164.5	c.747+1G>T	5:176656817	**NXF2**	ENST00000625106.3	c.233+2T>C	23:102317154
PAMR1	ENST00000611014.4	c.714+5G>C	11:35434507	**GPX5**	ENST00000469384.1	c.242-3C>T	6:28532318
CACNA1B	ENST00000371363.5	c.2093-3C>A	9:138010007				
ATXN2L	ENST00000570200.5	c.616+79G>T	16:28826469	**OPRM1**	ENST00000520708.5	c.864+1405G>A	6:154092877
LZTR1	ENST00000646124.1	c.1785+21A>G	22:20994748	**PTPRT**	ENST00000373198.8	c.3771+1049T>G	20:42101075
CES1	ENST00000360526.7	c.1318+724G>C	16:55809793	**ELP4**	ENST00000640954.1	c.259+1970A>G	11:31522061
CELF4	ENST00000591282.5	c.801+93C>A	18:37274218	**NEB**	ENST00000604864.5	c.6915+1336A>G	2:151652656
DDR1	ENST00000376569.7	c.418-287G>T	6:30890686	**GABRB2**	ENST00000393959.6	c.353-1057G>T	5:161337826
KRT17	ENST00000311208.12	c.433-195T>C	17:41623227	**COG6**	ENST00000455146.7	c.789-1074C>T	13:39686429
DGKI	ENST00000453654.6	c.1403-841A>T	7:137488530	**CPLANE1**	ENST00000508244.5	c.938+1393T>A	5:37237464
GADD45GIP1	ENST00000316939.2	c.350+644G>C	19:12956219	**ABCA8**	ENST00000586539.5	c.2765-1855T>G	17:68896868
CACNA1G	ENST00000515165.5	c.1924+645G>A	17:50576971	**AC093668.1**	ENST00000514917.2	c.425+1352A>G	7:102568483
CALD1	ENST00000361901.6	c.1171-428G>C	7:134957641	**ATP5MF-PTCD1**	ENST00000413834.5	c.121+1825T>A	7:99458261

Tabla 16. *Variantes seleccionadas de los ficheros generados por el script del* ***Anexo I*** *(Parte 2 de 2).*

Gen	Transcrito	Anotación	Posición cromosómica	Gen	Transcrito	Anotación	Posición cromosómica
NTRK3	ENST00000394480.6	c.2133+614A>T	15:87928577	**TREML4**	ENST00000341495.6	c.507-1699A>T	6:41234787
COL25A1	ENST00000399132.5	c.1020+571C>T	4:108883607	**ZMYND8**	ENST00000360911.7	c.3272-1970T>A	20:47212867
RTTN	ENST00000255674.11	c.5824-590G>A	18:70025438	**BCAP29**	ENST00000379119.6	c.589+1714G>T	7:107602219
RBCK1	ENST00000356286.9	c.1452+35C>T	20:429129	**VAV3**	ENST00000370056.8	c.321+1896C>T	1:107873005
RAB11A	ENST00000261890.6	c.41-1373A>G	15:65875959				
PIK3CD	ENST00000377346.8	c.2808C>T	1:9724365	**EPAS1**	ENST00000263734.4	c.2401C>A	2:46382538
NCKAP1	ENST00000360982.2	c.253C>T	2:183003310	**PGF**	ENST00000238607.10	c.150C>T	14:74949519
CHEK2	ENST00000404276.6	c.1116C>T	22:28695853	**NLRP1**	ENST00000354411.7	c.1308G>A	17:5559388
OR10H2	ENST00000305899.4	c.612T>C	19:15728655	**NLRX1**	ENST00000409109.5	c.2627G>T	11:119183138
TP53	ENST00000445888.6	c.1024C>T	17:7670685	**LMO7**	ENST00000321797.12	c.3281G>A	13:75841932
ELMO2	ENST00000290246.10	c.1684C>T	20:46371588	**ADAM17**	ENST00000310823.8	c.359T>C	2:9536700
KRAS	ENST00000557334.5	c.38G>A	12:25245347	**TYW1**	ENST00000359626.9	c.270G>A	7:66998951
KCNJ11	ENST00000339994.4	c.1054G>A	11:17387038	**JAK2**	ENST00000381652.3	c.1849G>T	9:5073770
SORL1	ENST00000260197.11	c.5914C>G	11:121621088	**SMARCA4**	ENST00000413806.7	c.2843A>G	19:11021759
BRCA1	ENST00000352993.7	c.2106C>G	17:43045738	**KRAS**	ENST00000256078.8	c.34G>C	12:25245351
TP53	ENST00000510385.5	c.92A>G	17:7675124	**KRAS**	ENST00000557334.5	c.35G>A	12:25245350
CLDN12	ENST00000394605.2	c.*581T>C	7:90413992	**CLIP4**	ENST00000401617.6	c.601G>A	2:29145268
KDR	ENST00000263923.5	c.802C>T	4:55113478	**TP53**	ENST00000619186.4	c.266G>A	17:7674220
CHRNA7	ENST00000306901.8	c.961G>T	15:32163306	**ARID1B**	ENST00000635849.1	c.2262G>A	6:157201166
KNG1	ENST00000447445.1	c.1100A>G	3:186743709				

ANEXO VI. Estudio de las mutaciones según la clasificación de Wimmer *et al*. 2007.

Tabla 17. *Tipo de mutación para los predictores ESEfinder y Human Splicing Finder (Parte 1 de 2).*

Cambio	Tipo de mutación según *ESEfinder*	Tipo de mutación según *Human Splicing Finder*	Cambio	Tipo de mutación según *ESEfinder*	Tipo de mutación según *Human Splicing Finder*
ATM c.2921+1G>A	Tipo IV	Tipo I	**CACNA1B c.2093-3C>A**	Tipo I	No efecto
NF1 c.1845+1G>A	Tipo I	Tipo I	**DIDO1 c.2214+2T>C**	Tipo I	Tipo I
COL5A1 c.925-2A>G	Tipo I	Tipo I	**ARHGAP28 c.955-1G>T**	Tipo I	Tipo IV
OXCT1 c.1248+5G>A	Tipo I	Tipo I	**FRMPD4 c.813+3C>G**	Tipo I	Tipo IV
NF1 c.288+1137G>A	Tipo II	Tipo II	**NIN c.1546-2A>G**	Tipo I	Tipo I
CFTR c.3718-2477C>T	Tipo II	Tipo II	**CLASRP C.100-1G>A**	Tipo I	Tipo I
AR c.2450-118A>G	Tipo II (*donor*)	Tipo II (*acceptor*)	**CHL1 c.197+1G>A**	Tipo I	Tipo I
GLA c.639+919G>A	Tipo II	Tipo II	**NPR3 c.892+1G>T**	Tipo IV	Tipo IV
COL2A1 c.192G>A	Tipo V	Tipo V	**MFSD8 c.554-1G>T**	Tipo I	Tipo I
ACADM c.382C>T	Tipo V	No efecto	**SEMA4A c.685+1G>A**	Tipo I	Tipo I
NF1 c.3362A>G	Tipo V	Tipo III	**GFPT2 c.2004+4C>T**	Tipo I	Tipo I
BRCA1 c.4484G>T	Tipo V	Tipo V	**PTPRT c.1865+1G>T**	Tipo I	Tipo I
CACNA1C c.1218-2A>G	Tipo I	Tipo I	**STK11 c.465-1G>T**	Tipo I	Tipo IV
GIT1 c.405+5G>C	Tipo I	Tipo IV	**NXF2 c.233+2T>C**	Tipo I	Tipo I
CHD4 c.3820-1G>T	Tipo I	No efecto	**GPX5 c.242-3C>T**	Tipo IV	No efecto
PIR c.273+1G>A	Tipo I	Tipo I	**ATXN2L c.616+79G>T**	No efecto	No efecto
UEVLD c.821-1G>C	No efecto	Tipo I	**LZTR1 c.1785+21A>G**	No efecto	No efecto
CTCFL c.755-1G>T	Tipo I	Tipo I	**CES1 c.1318+724G>C**	Tipo II	Tipo II
NF2 c.1488-1G>C	Tipo I	Tipo IV	**CELF4 c.801+93C>A**	No efecto	Tipo II
U2AF1 c.44+1G>A	Tipo I	Tipo I	**DDR1 c.418-287G>T**	No efecto	Tipo II
MAP4K4 c.640-1G>T	Tipo I	Tipo IV	**KRT17 c.433-195T>C**	Tipo II	Tipo II
RHCE c.939+1G>T	Tipo I	Tipo I	**DGKI c.1403-841A>T**	Tipo II	Tipo II
TP53 c.375+1G>A	Tipo I	Tipo I	**GADD45GIP c.350+644G>C**	No efecto	No efecto
KLHL2 c.1480+2T>C	Tipo I	Tipo I	**CACNA1G c.1924+645G>A**	No efecto	Tipo II
TSPAN17 c.747+1G>T	Tipo I	Tipo I	**CALD1 c.1171-428G>C**	No efecto	Tipo II

Tabla 17. *Tipo de mutación para los predictores ESEfinder y Human Splicing Finder (Parte 2 de 2).*

Cambio	Tipo de mutación según *ESEfinder*	Tipo de mutación según *Human Splicing Finder*	Cambio	Tipo de mutación según *ESEfinder*	Tipo de mutación según *Human Splicing Finder*
PAMR1 c.714+5G>C	Tipo I	Tipo IV	**OPRM1 c.864+1405G>A**	No efecto	Tipo II
PTPRT c.3771+1049T>G	No efecto	Tipo II	**KRAS c.38G>A**	Tipo III	Tipo III
ELP4 c.259+1970A>G	Tipo II	Tipo II	**KCNJ11 c.1054G>A**	Tipo V	No efecto
NEB c.6915+1336A>G	No efecto	Tipo II	**SORL1 c.5914C>G**	Tipo III	Tipo III
GABRB2 c.353-1057G>T	Tipo II	Tipo II	**BRCA1 c.2106C>G**	Tipo III	Tipo V
COG6 c.789-1074C>T	No efecto	No efecto	**TP53 c.92A>G**	Tipo III	Tipo V
CPLANE1 c.938+1393T>A	No efecto	No efecto	**CLDN12 c.*581T>C**	Tipo III	No se puede analizar
ABCA8 c.2765-1855T>G	No efecto	Tipo II	**KDR c.802C>T**	Tipo V	Tipo V
AC093668.1 c.425+1352A>G	No efecto	No efecto	**CHRNA7 c.961G>T**	Tipo V	Tipo V
ATP5MF-PTCD1 c.121+1825T>A	No efecto	Tipo II	**KNG1 c.1100A>G**	No efecto	No efecto
NTRK3 c.2133+614A>T	No efecto	No efecto	**EPAS1 c.2401C>A**	Tipo V	No efecto
COL25A1 c.1020+571C>T	No efecto	No efecto	**PGF c.150C>T**	Tipo V	Tipo V
RTTN c.5824-590G>A	No efecto	Tipo II	**NLRP1 c.1308G>A**	Tipo III	Tipo III
RBCK1 c.1452+35C>T	No efecto	No efecto	**NLRX1 c.2627G>T**	Tipo III	Tipo V
RAB11A c.41-1373A>G	No efecto	Tipo II	**LMO7 c.3281G>A**	No efecto	No efecto
TREML4 c.507-1699A>T	Tipo II	No efecto	**ADAM17 c.359T>C**	Tipo III	Tipo III
ZMYND8 c.3272-1970T>A	Tipo II	Tipo II	**TYW1 c.270G>A**	Tipo V	No efecto
BCAP29 c.589+1714G>T	No efecto	No efecto	**JAK2 c.1849G>T**	Tipo V	Tipo V
VAV3 c.321+1896C>T	No efecto	No efecto	**SMARCA4 c.2843A>G**	Tipo III	Tipo V
PIK3CD c.2808C>T	Tipo V	Tipo V	**KRAS c.34G>C**	Tipo V	No efecto
NCKAP1 c.253C>T	No efecto	Tipo V	**KRAS c.35G>A**	Tipo V	No efecto
CHEK2 c.1116C>T	No efecto	No efecto	**CLIP4 c.601G>A**	No efecto	No efecto
OR10H2 c.612T>C	No efecto	No efecto	**TP53 c.266G>A**	Tipo V	Tipo III
TP53 c.1024C>T	No efecto	No efecto	**ARID1B c.2262G>A**	Tipo V	Tipo III
ELMO2 c.1684C>T	Tipo V	Tipo III			

Anexo VII. Tipo de mutación según la clasificación de Wimmer et al. 2007 para los predictores *ESEfinder, Human Splicing Finder, NetGene2 y NNSplice.*

Tabla 18. *Tipo de mutación para los predictores ESEfinder, Human Splicing Finder, NetGene2 y NNSplice (Parte 1 de 4).*

Cambio	Tipo de mutación según ESEfinder	Tipo de mutación según Human Splicing Finder	Tipo de mutación según NetGene2	Tipo de mutación según NNSplice
ATM c.2921+1G>A	Tipo I/IV	Tipo I/IV	No efecto	Tipo I/IV
NF1 c.1845+1G>A	Tipo I/IV	Tipo I/IV	No efecto	Tipo I/IV
COL5A1 c.925-2A>G	Tipo I/IV	Tipo I/IV	No efecto	Tipo I/IV
OXCT1 c.1248+5G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
NF1 c.288+1137G>A	Tipo II	Tipo II	Tipo II	Tipo II
CFTR c.3718-2477C>T	Tipo II	Tipo II	Tipo II	Tipo II
AR c.2450-118A>G	Tipo II (donor)	Tipo II (acceptor)	Tipo II (donor)	Tipo II (donor)
GLA c.639+919G>A	Tipo II	Tipo II	Tipo II	No efecto
COL2A1 c.192G>A	Tipo III/V	Tipo III/V	No efecto	No efecto
ACADM c.382C>T	Tipo III/V	No efecto	No efecto	No efecto
NF1 c.3362A>G	Tipo III/V	Tipo III/V	No efecto	No efecto
BRCA1 c.4484G>T	Tipo III/V	Tipo III/V	No efecto	No efecto
CACNA1C c.1218-2A>G	Tipo I/IV	Tipo I/IV	No efecto	Tipo I/IV
GIT1 c.405+5G>C	Tipo I/IV	Tipo I/IV	No efecto	Tipo I/IV
CHD4 c.3820-1G>T	Tipo I/IV	No efecto	Tipo I/IV	Tipo I/IV
PIR c.273+1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
UEVLD c.821-1G>C	No efecto	Tipo I/IV	Tipo I/IV	No efecto
CTCFL c.755-1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
NF2 c.1488-1G>C	Tipo I/IV	Tipo I/IV	Tipo I/IV	No efecto
U2AF1 c.44+1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
MAP4K4 c.640-1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
RHCE c.939+1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
TP53 c.375+1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
KLHL2 c.1480+2T>C	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
TSPAN17 c.747+1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
PAMR1 c.714+5G>C	Tipo I/IV	Tipo I/IV	No efecto	No efecto

Tabla 18. *Tipo de mutación para los predictores ESEfinder, Human Splicing Finder, NetGene2 y NNSplice (Parte 2 de 4).*

Cambio	Tipo de mutación según ESEfinder	Tipo de mutación según Human Splicing Finder	Tipo de mutación según NetGene2	Tipo de mutación según NNSplice
CACNA1B c.2093-3C>A	Tipo I/IV	No efecto	No efecto	No efecto
DIDO1 c.2214+2T>C	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
ARHGAP28 c.955-1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
FRMPD4 c.813+3C>G	Tipo I/IV	Tipo I/IV	No efecto	No efecto
NIN c.1546-2A>G	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
CLASRP C.100-1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
CHL1 c.197+1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
NPR3 c.892+1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
MFSD8 c.554-1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	No efecto
SEMA4A c.685+1G>A	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
GFPT2 c.2004+4C>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
PTPRT c.1865+1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
STK11 c.465-1G>T	Tipo I/IV	Tipo I/IV	Tipo I/IV	No efecto
NXF2 c.233+2T>C	Tipo I/IV	Tipo I/IV	Tipo I/IV	Tipo I/IV
GPX5 c.242-3C>T	Tipo I/IV	No efecto	Tipo I/IV	Tipo I/IV
ATXN2L c.616+79G>T	No efecto	No efecto	No efecto	No efecto
LZTR1 c.1785+21A>G	No efecto	No efecto	No efecto	No efecto
CES1 c.1318+724G>C	Tipo II	Tipo II	No efecto	No efecto
CELF4 c.801+93C>A	No efecto	Tipo II	No efecto	No efecto
DDR1 c.418-287G>T	No efecto	Tipo II	No efecto	No efecto
KRT17 c.433-195T>C	Tipo II	Tipo II	No efecto	No efecto
DGKI c.1403-841A>T	Tipo II	Tipo II	Tipo II	Tipo II
GADD45GIP c.350+644G>C	No efecto	No efecto	No efecto	No efecto
CACNA1G c.1924+645G>A	No efecto	Tipo II	Tipo II	No efecto
CALD1 c.1171-428G>C	No efecto	Tipo II	No efecto	No efecto
OPRM1 c.864+1405G>A	No efecto	Tipo II	No efecto	No efecto
PTPRT c.3771+1049T>G	No efecto	Tipo II	Tipo II	No efecto
ELP4 c.259+1970A>G	Tipo II	Tipo II	No efecto	No efecto
NEB c.6915+1336A>G	No efecto	Tipo II	No efecto	No efecto

Tabla 18. *Tipo de mutación para los predictores ESEfinder, Human Splicing Finder, NetGene2 y NNSplice (Parte 3 de 4).*

Cambio	Tipo de mutación según ESEfinder	Tipo de mutación según Human Splicing Finder	Tipo de mutación según NetGene2	Tipo de mutación según NNSplice
GABRB2 c.353-1057G>T	Tipo II	Tipo II	No efecto	No efecto
COG6 c.789-1074C>T	No efecto	No efecto	No efecto	Tipo II
CPLANE1 c.938+1393T>A	No efecto	No efecto	No efecto	No efecto
ABCA8 c.2765-1855T>G	No efecto	Tipo II	No efecto	No efecto
AC093668.1 c.425+1352A>G	No efecto	No efecto	Tipo II	No efecto
ATP5MF-PTCD1 c.121+1825T>A	No efecto	Tipo II	No efecto	No efecto
NTRK3 c.2133+614A>T	No efecto	No efecto	No efecto	No efecto
COL25A1 c.1020+571C>T	No efecto	No efecto	No efecto	No efecto
RTTN c.5824-590G>A	No efecto	Tipo II	No efecto	Tipo II
RBCK1 c.1452+35C>T	No efecto	No efecto	No efecto	No efecto
RAB11A c.41-1373A>G	No efecto	Tipo II	No efecto	No efecto
TREML4 c.507-1699A>T	Tipo II	No efecto	No efecto	No efecto
ZMYND8 c.3272-1970T>A	Tipo II	Tipo II	No efecto	No efecto
BCAP29 c.589+1714G>T	No efecto	No efecto	No efecto	No efecto
VAV3 c.321+1896C>T	No efecto	No efecto	No efecto	No efecto
PIK3CD c.2808C>T	Tipo III/V	Tipo III/V	No efecto	No efecto
NCKAP1 c.253C>T	No efecto	Tipo III/V	No efecto	No efecto
CHEK2 c.1116C>T	No efecto	No efecto	No efecto	No efecto
OR10H2 c.612T>C	No efecto	No efecto	No efecto	No efecto
TP53 c.1024C>T	No efecto	No efecto	Tipo III/V	No efecto
ELMO2 c.1684C>T	Tipo III/V	Tipo III/V	Tipo III/V	No efecto
KRAS c.38G>A	Tipo III/V	Tipo III/V	Tipo III/V	No efecto
KCNJ11 c.1054G>A	Tipo III/V	No efecto	No efecto	No efecto
SORL1 c.5914C>G	Tipo III/V	Tipo III/V	No efecto	No efecto
BRCA1 c.2106C>G	Tipo III/V	Tipo III/V	No efecto	No efecto
TP53 c.92A>G	Tipo III/V	Tipo III/V	No efecto	No efecto
CLDN12 c.*581T>C	Tipo III/V	No se puede analizar	No efecto	No efecto
KDR c.802C>T	Tipo III/V	Tipo III/V	Tipo III/V	No efecto

Tabla 18. *Tipo de mutación para los predictores ESEfinder, Human Splicing Finder, NetGene2 y NNSplice (Parte 4 de 4).*

Cambio	Tipo de mutación según ESEfinder	Tipo de mutación según Human Splicing Finder	Tipo de mutación según NetGene2	Tipo de mutación según NNSplice
CHRNA7 c.961G>T	Tipo III/V	Tipo III/V	No efecto	No efecto
KNG1 c.1100A>G	No efecto	No efecto	No efecto	Tipo III/V
EPAS1 c.2401C>A	Tipo III/V	No efecto	No efecto	No efecto
PGF c.150C>T	Tipo III/V	Tipo III/V	No efecto	No efecto
NLRP1 c.1308G>A	Tipo III/V	Tipo III/V	Tipo III/V	No efecto
NLRX1 c.2627G>T	Tipo III/V	Tipo III/V	No efecto	No efecto
LMO7 c.3281G>A	No efecto	No efecto	No efecto	No efecto
ADAM17 c.359T>C	Tipo III/V	Tipo III/V	No efecto	No efecto
TYW1 c.270G>A	Tipo III/V	No efecto	Tipo III/V	No efecto
JAK2 c.1849G>T	Tipo III/V	Tipo III/V	No efecto	No efecto
SMARCA4 c.2843A>G	Tipo III/V	Tipo III/V	No efecto	No efecto
KRAS c.34G>C	Tipo III/V	No efecto	Tipo III/V	No efecto
KRAS c.35G>A	Tipo III/V	No efecto	Tipo III/V	No efecto
CLIP4 c.601G>A	No efecto	No efecto	Tipo III/V	No efecto
TP53 c.266G>A	Tipo III/V	Tipo III/V	No efecto	No efecto
ARID1B c.2262G>A	Tipo III/V	Tipo III/V	Tipo III/V	No efecto