

RESUMEN

Introducción: El manejo de los linfomas se ha visto optimizado gracias a la mejor comprensión de su biología y el desarrollo de herramientas terapéuticas novedosas. Este trabajo pretende determinar las vías moleculares desreguladas en linfomas frente a los nódulos linfáticos activados fisiológicamente. Para probar su implicación biológica tumoral, se diseña un modelo de *knockout* sobre un gen sobreexpresado en la célula linfomatosa.

Materiales y métodos: A través de la base de datos pública GEO-NCBI se obtienen los estudios *microarray* de 116 pacientes, analizándolos mediante un modelo estadístico de predicción de fenotipos. El *knockout* se diseña utilizando el sistema de alta precisión CRISPR/Cas, proponiéndose diferentes oligonucleótidos integrados en el plásmido pX330.

Resultados: Los linfomas presentan desregulación a nivel de las vías implicadas en la proliferación celular, así como en las mediadas por mensajeros extracelulares. Entre ellas, la señalización serotoninérgica se encuentra significativamente alterada, diseñándose un modelo de *knockout* sobre el gen *HTR3A*, sobreexpresado en linfomas foliculares ($P=0,004$).

Conclusiones: Los resultados obtenidos sientan la base estadística que justifica el desarrollo de un experimento de *knockout* sobre la vía serotoninérgica mediante el sistema CRISPR/Cas propuesto en este estudio. Determinar su implicación biológica puede esclarecer su potencial como diana terapéutica.

ABSTRACT

Introduction: The management of lymphomas has been enhanced due to a better understanding of their biology and the development of new therapeutic tools. The aim of this work is to determine dysregulated pathways in lymphomas in comparison with reactive lymph nodes. In order to probe their relevance in tumoral biology, a knockout model is designed disrupting an overexpressed gen in a lymphomatous cell.

Materials and methods: 116 microarray assays are obtained from the public database GEO-NCBI, using for their analysis a phenotype prediction statistic model. The high precision CRISPR/Cas system is used to design the knockout, proposing different oligonucleotides integrated in the plasmid pX330.

Results: Dysregulation of lymphomas involve cellular proliferation pathways as well as signaling regulated by extracellular messengers. Among them, serotonergic signaling is significantly altered. A knockout model is designed to disrupt *HTR3A* gene, overexpressed in follicular lymphomas ($P=0,004$).

Conclusions: These results establish the statistical base to justify the development of an experiment to knockout the serotonergic pathway using the CRISPR/Cas system proposed in this work. Determination of its biological implication could clarify its potential as therapeutic target.



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

FACULTAD DE MEDICINA Y CIENCIAS DE LA SALUD

TRABAJO DE FIN DE GRADO. MAYO 2021

GRADO EN MEDICINA

**DISEÑO DE KNOCKOUT MEDIANTE CRISPR/CAS PARA
GENES ALTERADOS EN LINFOMAS OBTENIDOS
MEDIANTE MINERÍA DE DATOS EN GEO-NCBI**

AUTOR: DANIEL MUNÁRRIZ LORCA

TUTORES: IVÁN FERNÁNDEZ VEGA, LUIS MANUEL QUIRÓS FERNÁNDEZ

ÍNDICE

ÍNDICE	2
1. INTRODUCCIÓN	3
2. ESTADO ACTUAL DEL TEMA	5
2.1. Minería de datos	5
2.2. Base de datos Geo-NCBI	5
2.3. Diseño de <i>knockout</i> mediante CRISPR/Cas	7
3. HIPÓTESIS, OBJETIVOS Y LIMITACIONES	9
3.1. Hipótesis	9
3.2. Objetivos	9
3.3. Limitaciones	9
4. MATERIALES Y MÉTODOS	10
4.1. Selección de casos	10
4.2. Procesamiento y análisis de los datos	10
4.3. Método para el diseño del <i>knockout</i>	12
4.4. Análisis estadístico	14
5. RESULTADOS	15
5.1. Análisis computacional de los <i>microarrays</i> de linfomas.	15
5.2. Diseño del sgRNA para el gen <i>HTR3A</i>	20
6. DISCUSIÓN	22
7. CONCLUSIONES	25
8. BIBLIOGRAFÍA	26
9. ANEXO 1	30

1. INTRODUCCIÓN

Los linfomas son un grupo amplio y heterogéneo de tumores sólidos que derivan de células sanguíneas. Actualmente representan aproximadamente el 5 % de la incidencia del cáncer y causan cerca del 4 % de todas las muertes por enfermedades malignas. La incidencia de linfoma ha aumentado espectacularmente en la última mitad del siglo XX afectando a ambos sexos y a todos los grupos de edad (1). Así, en Europa la incidencia estimada alcanza los 19 casos por cada 100.000 habitantes (2).

Se pueden organizar atendiendo a la estirpe celular que los compone, bien sea linfomas de células B o células T/natural killer (NK), aunque tradicionalmente esta enfermedad se ha dividido en dos grandes grupos: linfoma no Hodgkin (90% del total) y linfoma de Hodgkin (3). Los tipos de linfoma no Hodgkin más comunes son el linfoma difuso de células grandes de estirpe B (DLBCL) y el linfoma folicular (FL) y juntos representan la mitad de todos los casos de linfomas (4) (Figura 1).

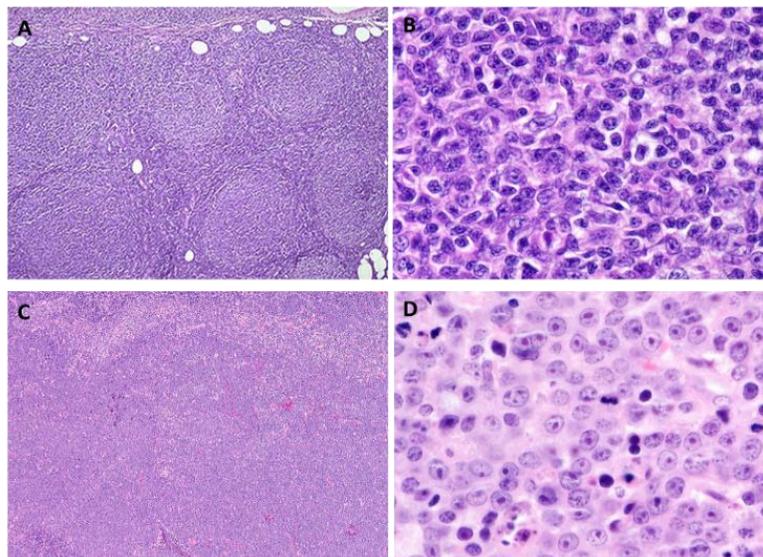


Figura 1: Imágenes histológicas de linfomas. A-B) Linfoma folicular. C-D) DLBCL. (Imágenes propias).

En los últimos años la inmunoterapia y fármacos contra dianas específicas han aumentado el abanico de abordajes terapéuticos de los que se dispone en la práctica clínica (anticuerpos monoclonales, linfocitos T-CAR) (5). Estos avances son el reflejo de la investigación dedicada a comprender la biología de las células linfomatosas, desde una perspectiva cada vez más molecular y genética. No obstante, estas neoplasias siguen constituyendo un reto terapéutico para el que resulta necesario aplicar técnicas vanguardistas que permitan implementar los resultados obtenidos. Estos beneficios deberán materializarse en un incremento de la calidad de vida y supervivencia de los pacientes.

En el presente proyecto se emplea la minería de datos con el fin de profundizar en las desregulaciones genéticas que acontecen en la célula linfomatososa. Así, a través del análisis estadístico basado en la predicción de fenotipos se pretende vislumbrar nuevas vías biológicas alteradas y proponer líneas de tratamiento novedosas. En un segundo apartado, se diseña un modelo de *knockout* con el que poder modular la expresión de un gen sobreexpresado con el fin de comprender los efectos derivados de su inhibición. Para este cometido se emplea el sistema CRISPR/Cas, herramienta de ingeniería genética galardonada con un premio Nobel, capaz de realizar modificaciones en el ADN con una alta precisión.

Por tanto, este trabajo persigue realizar una investigación transversal que abarque desde el análisis estadístico de la desregulación genética tumoral hasta el enfoque experimental con el que obtener nuevas dianas para la terapéutica oncológica futura.

2. ESTADO ACTUAL DEL TEMA

2.1. Minería de datos

La exploración de datos o también llamado minería de datos (del inglés *data mining*), es el proceso estadístico que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Se ejecuta aplicando diferentes técnicas informáticas *in silico* y ha experimentado en los últimos años un crecimiento exponencial en biomedicina orientada a identificar posibles biomarcadores de enfermedad (6, 7). Estas técnicas permiten analizar simultáneamente numerosos genes: chips *microarray* o secuenciación de ARN (RNA-seq).

Su objetivo general es la extracción de información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior, mediante análisis automático. Se pretende identificar grupos en los datos que luego pueden ser utilizados para obtener resultados más precisos de predicción por sistemas de soporte de decisiones.

2.2. Base de datos Geo-NCBI

Las técnicas de minería de datos se aplicaban principalmente sobre información contenida en almacenes de datos (8). En el campo del estudio del cáncer destacan sobre todo el proyecto TCGA (The Cancer Genome Atlas) y GEO (Gene Expression Omnibus).

GEO es una base de datos pública perteneciente al NCBI que almacena datos procedentes de *microarrays*, experimentos de *Next Generation Sequencing* y otros tipos de datos procedentes de genómica funcional enviados por la comunidad investigadora. Los datos almacenados en GEO cumplen con el estándar MIAME (9): requisitos sobre la información mínima que debería ser almacenada en experimentos *microarray*.

Los experimentos de *microarray* se basan en la diferente expresión que existe entre los genes de dos muestras : una problema y otra control. La primera suele provenir de tumores o células sometidas a un condicionante (como un tratamiento o un ambiente concreto). Se acepta que la expresión génica de una célula es reflejada por la cantidad de ARN mensajero (ARNm) que posee. Para analizarlo, este ARNm será depositado en chips, conteniendo cada una de sus celdas ADN monocatenario que represente un gen determinado. Como paso previo, el ARNm deberá ser transformado en ADN para que pueda hibridar con las hebras situadas en las diferentes celdas. Para ello se emplea la transcripción reversa cuyo resultado es el ADN complementario (ADNc) a las hebras de ARNm extraídas. Además, los nucleótidos empleados en la síntesis de la cadena están marcados con cianina fluorescente, reconocible por el láser que analiza los chips. El ADNc resultante de cada muestra poseerá cianina de un color diferente. Por convenio en los controles se emplea el verde (Cy3) y en los casos el rojo (Cy5). Una vez depositado el ADNc de ambas muestras en el chip, éste hibridará con la sonda monocatenaria complementaria. Posteriormente, cada celda es analizada en función de la longitud de onda emitida en su fluorescencia, estableciéndose un cociente (Cy5/Cy3). De esta forma, si la expresión génica de ambas muestras es igual, el cociente $Cy5/Cy3=1$ y será reflejado en amarillo, si es mayor en la muestra problema $Cy5/Cy3>1$ y tendrá color rojo y si es mayor en el control $Cy5/Cy3<1$, mostrándose en verde (10). Así, se puede comparar de manera rápida la expresión de miles de genes simultáneamente (Figura 2).

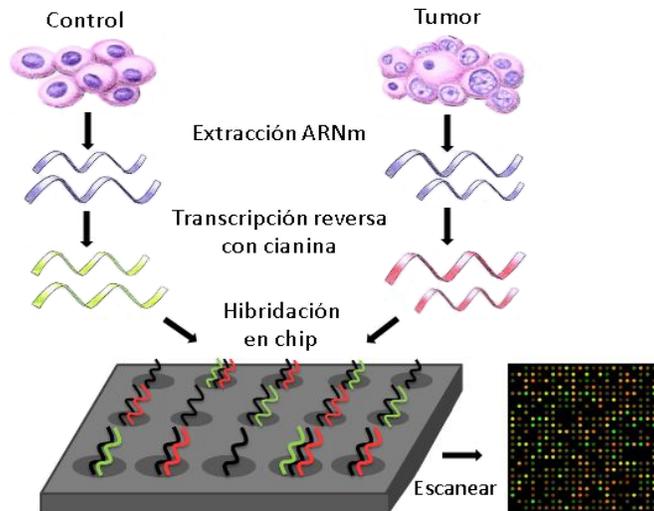


Figura 2: Fundamentos de un experimento microarray (Imagen propia).

2.3. Diseño de *knockout* mediante CRISPR/Cas

La desactivación genética (*knockout* en inglés) permite eliminar la expresión de un determinado gen en un organismo concreto. Así se puede analizar no solo las consecuencias de su desactivación, sino las asociaciones e implicaciones que acarrea en la expresión de otros genes (11). Para llevar a cabo el *knockout* existen numerosas técnicas, destacando las endonucleasas específicamente dirigidas. Son estas últimas las que en la actualidad han experimentado un mayor impulso, debido al desarrollo del sistema CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats). Es una técnica joven y en auge gracias a su versatilidad, descubriéndose de manera casi constante nuevos potenciales de uso. Así, Emmanuelle Charpentier y Jennifer Doudna fueron galardonadas con el premio Nobel de química 2020 por sus contribuciones al descubrimiento y desarrollo del CRISPR/Cas (12).

Su origen se encuentra en la naturaleza, formando parte del sistema inmunitario de las bacterias que les permite hacer frente a la infección por material genético extraño, como el de los virus bacteriófagos (13). El sistema CRISPR/Cas se compone de dos estructuras

básicas: una secuencia guía de ARN (single-guide RNA o sgRNA) y una endonucleasa (Cas). Este sgRNA presentará una secuencia de unos 20 nucleótidos que debe hibridar con el gen en el que se quiere realizar el corte, seguida en su extremo 3' de una región conocida como PAM (protospacer adjacent motif) que permitirá el reconocimiento y unión de la endonucleasa Cas (14). Debido a la variabilidad del genoma y la frecuencia de polimorfismos de un solo nucleótido (SNPs acrónimo inglés), cuando se diseña la secuencia de bases de un sgRNA debe buscarse que existan el menor número posible de discordancias (*mismatches* en inglés), especialmente en los 10-12 nucleótidos adyacentes a la PAM. Esta docena de bases recibe el nombre de región semilla (*seed region*) y en ella los *mismatches* son peor tolerados, pudiendo conducir incluso a la pérdida de función del sgRNA (15). El segundo componente de este sistema lo forma la endonucleasa Cas. Esta proteína ha sido aislada en numerosas bacterias, habiéndose extendido el uso de la Cas9, procedente del *Streptococcus pyogenes*. Así, Cas9 requiere una secuencia PAM formada por cualquier nucleótido seguido de dos guaninas (NGG). Una vez reconocida, se coloca de tal forma que es capaz de realizar un corte preciso a 3 nucleótidos *upstream* (en sentido 5') de la PAM (16) (Figura 3).

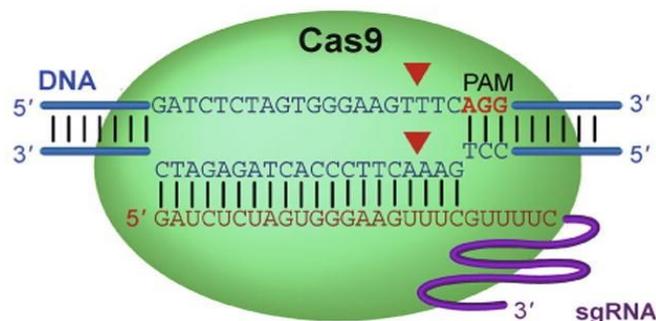


Figura 3: Representación del proceso de hibridación entre el sgRNA, la endonucleasa Cas9 y la secuencia objetivo. En la imagen se aprecia el corte a 3 nucleótidos de distancia de la secuencia PAM en sentido 5'. Imagen tomada de Wu X, Kriz AJ et al (17).

3. HIPÓTESIS, OBJETIVOS Y LIMITACIONES

3.1. Hipótesis

Se pretende analizar las diferencias en la expresión génica que existen entre las células linfomatosas y las células inmunitarias activadas de forma fisiológica en un nódulo linfático sano. Además, se busca diseñar *in silico* un experimento con el que testar qué efecto produce la eliminación de un gen significativamente sobreexpresado, identificado mediante análisis estadístico, sobre la biología de la célula linfomatososa.

3.2. Objetivos

1. Reconocer alteraciones en la expresión génica que puedan estar involucradas en la etiopatogenia y proliferación de la célula linfomatososa. Para ello, este proyecto se basa en el empleo de la minería de datos, así como el proceso estadístico de predicción de fenotipos. Las células inmunes reactivas a procesos no tumorales son tomadas como referencia de expresión génica fisiológica.

2. Establecer las vías biológicas o *pathways* alteradas en las que son partícipes los genes con una desregulación más significativa.

3. Diseño computacional de un modelo *knockout* utilizando la metodología de edición génica CRISPR/Cas para un gen concreto.

3.3. Limitaciones

Debido a la situación actual, con presencialidad y movilidad limitadas, el proyecto debe basarse en modelos, sin llevar a la práctica experimental real la modificación genética.

4. MATERIALES Y MÉTODOS

4.1. Selección de casos

En el presente estudio se emplearon datos de expresión génica de 116 ensayos con chips de *microarray* procedentes de la plataforma pública GEO DataSets-NCBI, accesibles a través del código de acceso GEO: GSE23647. Las 116 muestras proceden de pacientes con al menos una adenopatía, pudiendo deberse a nódulos linfáticos reactivos (utilizados aquí como controles, reflejando la expresión génica en un nódulo linfático fisiológicamente activado) o linfomatosos. Así 23 de las adenopatías son reactivas, 35 linfomas foliculares, 19 linfomas B difusos de célula grande (DLBCL), 19 linfomas de Hodgkin y otras 20 muestras con diversos linfomas (del manto, zona marginal, célula T, anaplásico, linfoma de célula pequeña y Burkitt). Los chips de *microarray* empleados en el análisis proceden de la compañía *Affimetrix*[®], siendo analizados cerca de 20000 genes por ensayo, generándose una matriz A de 2×10^4 filas por 116 columnas.

4.2. Procesamiento y análisis de los datos

Para el procesamiento y análisis de los datos se llevó a cabo una colaboración científica con un equipo multidisciplinar del área de matemática aplicada de la Universidad de Oviedo, a través de Juan Luis Fernández Martínez, catedrático y director del grupo de problemas inversos, optimización y aprendizaje automático, así como con Juan Ron Cosmen, biólogo, y Enrinque Juan de Andrés Galiana, doctor en informática. Para el tratamiento de los datos se emplearon R software 4.0.2 y el entorno de desarrollo integrado R Studio. Además, se optimizó R con paquetes estadísticos de *Bioconductor*, diseñado

específicamente para el manejo de resultados procedentes de experimentos *microarray* (18). Para el análisis de las vías biológicas en concreto se empleó el paquete *Reactome pathways analysis* (19).

Con objeto de obtener la firma genética que pueda predecir con mayor fiabilidad la clase de una muestra dada, deben seleccionarse los genes más discriminatorios, con una desregulación más significativa. Para ello, se puede atender al cociente de expresión entre dos clases (p.ej. enfermedad/control), también conocido como *Fold change* (FC). De manera habitual el cociente obtenido es expresado en forma de logaritmo en base 2, con objeto de facilitar el manejo de los valores. Así, los genes con el valor absoluto de su cociente ($|FC|$) mayor son aquellos con una desregulación más marcada frente al control. Para caracterizar cuáles son los genes más discriminatorios, se emplea el ratio de Fisher (FR). Este parámetro es directamente proporcional a la diferencia de expresión frente al control e inversamente proporcional a la varianza (σ^2) y, por tanto, es sensible a la dispersión de los valores. Los genes son ordenados de forma decreciente en función de su FR, obteniéndose los genes más discriminatorios (20). Mediante este procesamiento se reduce de forma drástica el número inicial de los cerca de 20000 genes, la mayoría de los cuales no son significativos.

Posteriormente, se busca la firma genética mínima capaz de predecir la clase a la que pertenezca una muestra dada. Con este fin, es empleado un algoritmo de optimización. Éste desarrolla diferentes modelos con los que poner a prueba la capacidad discriminatoria de cada firma genética testada, cuya estimación de precisión está basada en la validación cruzada: *Leave-One-Out-Cross-Validation* (LOOCV). Finalmente, también se busca la vía biológica alterada en la que se relacionan y ubican los genes más discriminatorios (20).

4.3. Método para el diseño del *knockout*

Para el silenciamiento del gen elegido, se propone un experimento de *knockout* mediante el sistema CRISPR/Cas basado en unión de extremos no homólogos. Las secuencias genéticas son obtenidas a través de la plataforma pública Gene-NCBI. En ella se encuentran recogidas los diferentes genes de *Homo Sapiens* en forma de pares de bases, así como las regiones codificantes, intrones, *locus*, variaciones, etc.

Una vez seleccionado el gen objetivo, su secuencia es manejada mediante programas bioinformáticos. Para el presente proyecto se elige el software *SnapGene*[®]. Así, el primer paso tras la lectura de la secuencia genética es la determinación de los exones que comparten las diferentes isoformas de la proteína codificada. Esto es debido a que, de no escoger una secuencia codificante del exón común a todas las isoformas, se corre el riesgo de que algunas de éstas sigan siendo viables y el experimento CRISPR/Cas pierda eficacia.

Una vez localizada una secuencia común se procede al diseño del sgRNA. Para la optimización de este proceso se emplea el programa CRISPOR (21). Así, de entre los diferentes sgRNA posibles ha de escogerse el más adecuado para cada experimento. El programa ofrece una serie de puntuación para cada sgRNA en función de diferentes escalas. Se escoge el sgRNA atendiendo a los SNPs que existan para cada secuencia, al *MIT Specificity Score* (mide especificidad), *Doench'16* (eficiencia) *Outcome* (predice los resultados tras la rotura en ambas hebras de ADN) y los *off-target events* (cortes en otras posiciones del ADN) en función de los *mismatches* entre el sgRNA y otras secuencias del genoma. Este último parámetro es de suma importancia cuando el experimento tiene una finalidad *in vivo*,

debido a los riesgos potenciales de generar mutaciones indeseadas en el organismo tratado. Por ello, se debe prestar atención a qué genes podrían afectarse y la importancia biológica que éstos tienen. Además, el sgRNA debe ser también escogido en función de la secuencia PAM, que ha de ser compatible con la endonucleasa Cas empleada. Para el presente estudio se elige la Cas9 por la simplicidad de su PAM. Por ello, todos los sgRNA elegidos han de continuarse con la secuencia NGG. Además, el empleo de múltiples sgRNA mejora la eficacia del experimento, por ello se diseñan 3 sgRNA diferentes para este estudio (22). Tras este proceso se procede a la determinación de los oligonucleótidos que deben ser producidos.

Para este proyecto se proponen dos opciones de administración del sistema CRISPR/Cas: RNPs y vector plasmídico introducidos mediante electroporación. Como vector se escoge el plásmido pX330 (Figura 4), internacionalmente utilizado para este tipo de experimentos (23) y disponible en el repositorio sin ánimo de lucro *AddGene*. PX330 codifica el esqueleto del sgRNA y la proteína Cas9, así como un gen de resistencia a la ampicilina. Además, contiene el promotor U6, que requiere una guanina en la posición +1 de la secuencia del sgRNA que, si éste no la presenta en sus 20 nucleótidos, debe ser añadida (24). Para la digestión de este plásmido se utiliza el enzima de restricción BbsI, generando extremos cohesivos, no homólogos, en la posición 245 (5'-GGTG-3') y 267 (5'-GTTT-3') del plásmido (25). Por ello, los oligonucleótidos producidos han de contener en sus extremos las secuencias 5'-CACC-3' Y 5'-AAAC-3' con el fin de obtener la hibridación con el plásmido previamente digerido. El proceso de modificación plasmídica podría ser realizado en el laboratorio, pero atendiendo a la eficiencia se recomienda pedir su fabricación específica.

De la misma forma, una vez diseñado el sgRNA se encarga su producción por una entidad externa.

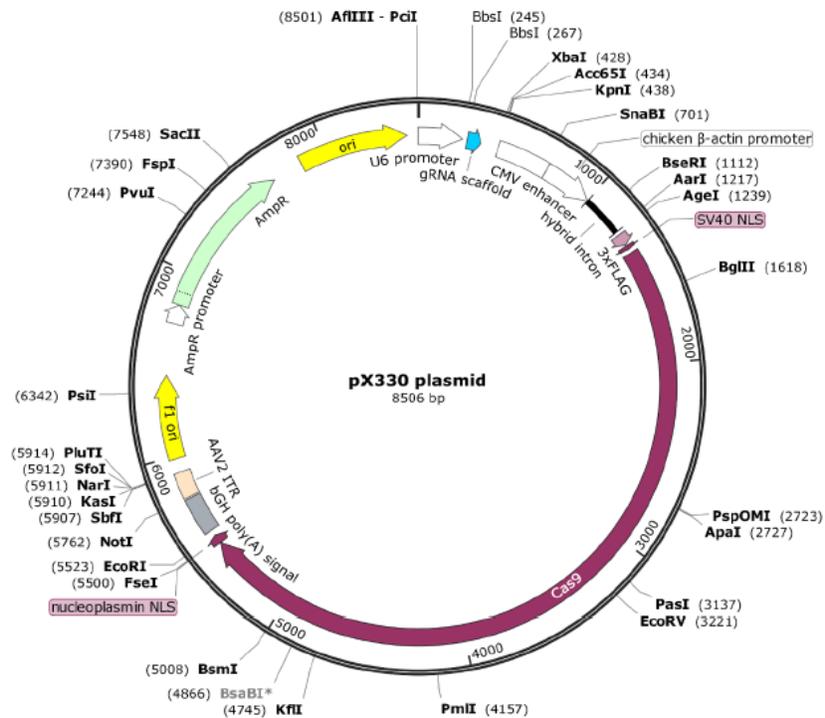


Figura 4: Plásmido pX330. En el interior se muestra el nombre de los genes contenidos. Excéntricamente se observan los nombres de los enzimas de restricción capaces de actuar sobre este plásmido, así como la localización de su secuencia diana.

Una vez se recibe el plásmido pX330 modificado y/o los oligonucleótidos de sgRNA, se procede a la realización del experimento *knockout* mediante su introducción en las células linfomatosas de cultivo (fase experimental).

4.4. Análisis estadístico

Para el análisis inferencial, se compararon los resultados de cada variable cuantitativa para cada grupo utilizando la prueba T de Student para variables independientes. El análisis estadístico fue realizado empleando el software R 4.0.2 para Windows, con un valor de significación de $P < 0,05$.

5. RESULTADOS

1.1 Análisis computacional de los *microarrays* de linfomas.

En primer lugar, se analizaron los linfomas en su conjunto (formando una sola clase o fenotipo) frente a los nódulos linfáticos reactivos. Como resultado se obtuvieron 55 vías biológicas alteradas significativamente (Tabla 1). Las 3 con mayor significación fueron la señalización de receptores tirosina-quinasa ($P=1,63 \cdot 10^{-6}$), organización de la matriz extracelular ($P=4,43 \cdot 10^{-5}$) y señalización mediada por interleucinas ($P=6,61 \cdot 10^{-5}$).

Tabla 1: Listado de las 55 vías biológicas alteradas significativamente, ordenadas de forma decreciente según su significación. Se observa cómo las vías de señalización (extracelular/intracelular) constituyen uno de los pathways más frecuentemente alterados.

VÍA BIOLÓGICA	P VALOR	VÍA BIOLÓGICA	P VALOR
Señalización vía receptor tirosina quinasa	1,63·10 ⁻⁶	Integración del metabolism energético	0,01986809
Organización de la matriz extracelular	4,43·10 ⁻⁵	Ciclo de liberación de serotonina	0,01986809
Señalización mediada por interleucinas	6,61·10 ⁻⁵	Ciclo de liberación de noradrenalina	0,01986809
Señalización intracelular por segundos mensajeros	0,00011518	Fase mitótica G1-G1/S	0,01986809
Enfermedades de trasducción de la señalización	0,00132086	Dianas ERK/MAPK	0,01986809
Sistema neuronal	0,00132086	Inactivación de ERKs	0,02142765
Activación plaquetaria, señalización y agregación	0,00132086	Señalización por receptor de complejo TGF-β	0,02147693
Transmisión por sinapsis química	0,00135775	Respuestas celulares a estímulos externos	0,02243778
Señalización AKT activada por PIP3	0,00213376	Cascadas de señalización MAPK	0,022567
Señalización por TGF-β	0,00215776	Respuesta a Ca ²⁺ citosólico plaquetario elevado	0,02322508
Ciclo de liberación de transmisores	0,00225474	Sensibilización plaquetaria por LDL	0,02648992
Homeostasis plaquetaria	0,00338045	Degradación mediada por APC/C de proteínas del ciclo celular	0,02725872
Formación de colágeno	0,00432982	Regulación del ciclo mitótico celular	0,02725872
Señalización FLT3	0,00638835	Eventos calcio dependientes	0,02923579
Señalización familia IL-1	0,00638835	Replicación ADN	0,02942049
Respuesta a infecciones	0,00638835	Regulación PTEN	0,03085872
Señalización MAPK1/MAPK3	0,00721547	Señalización por interleucina-10	0,03085872
Respuesta a infección VIH	0,00815201	Ciclo de liberación del neurotransmisor glutamato	0,03085872
Respuesta a estrés celular	0,00855839	Activación NF-kB mediada por TRAF6	0,03085872
Señalización por NTRK1 (TRKA)	0,0101003	Degranulación plaquetaria	0,03085872
Cascada RAF/MAP quinasa	0,01130296	Señalización por VEGF	0,03183485
Ensamblaje de fibrillas colágenas	0,01207851	Cambio a estado post-replicativo	0,03427541
Ciclo de liberación de dopamina	0,01545996	Señalización mediada por Receptor célula B (BCR)	0,03427541
Señalización de moléculas de adhesión neuronal	0,01545996	Señalización de SCF-KIT	0,03805089
Regulación de p53	0,01545996	Eventos mediados por proteína G	0,0385653
Transición G1/S	0,01785322	Transmisión por receptores de neurotransmisores y señalización postsináptica	0,04119099

Posteriormente, se analizaron los linfomas difusos de célula grande frente a los nódulos linfáticos reactivos. Se encontraron 33 vías con una alteración significativa (Tabla 2). Las 3 con una significación mayor son la relacionada con la organización de la matriz extracelular ($P=0,0041$), la señalización a través de la vía FLT3 ($P=0,0043$) y el conjunto de genes relacionados con la activación, señalización y agregación plaquetaria ($P=0,0059$).

Tabla 2: Listado de las 33 vías biológicas alteradas significativamente en DLBCL, ordenadas de forma decreciente según su significación. La desregulación de la vía MAPK aparece con alta frecuencia.

VÍA BIOLÓGICA	P VALOR	VÍA BIOLÓGICA	P VALOR
Organización de la matriz extracelular	0,004060041	Señalización por interleucinas	0,022444743
Señalización vía FLT3	0,004387946	Ensamblaje de las cadenas colágenas y otras estructuras multiméricas	0,028995906
Activación plaquetaria, señalización y agregación	0,005974059	Señalización NCAM para crecimiento de dendritas	0,032747541
Señalización MAPK1/MAPK3	0,008560553	Respuesta celular a stress térmico	0,033110312
Cascada RAF/MAP quinasa	0,008560553	Sistema neuronal	0,036455737
Señalización intracelular por segundos mensajeros	0,008560553	Regulación de la expresión génica vía NR1H3 & NR1H2 ligada al transporte y flujo de colesterol	0,038279854
Señalización por receptores tirosín quinasa	0,008560553	Factores de interacción con VIH	0,038279854
Respuesta celular a stress	0,008560553	Señalización AKT activada por PIP3	0,038279854
Desórdenes de transducción de señal	0,008560553	Regulación de la respuesta al calor mediada por HSF1	0,038279854
Transmisión a través de sinapsis química	0,011017978	Eventos relacionados con receptores sensibles a insulina (IRS) mediados por IGF-1	0,038279854
Transcripción de genes reguladores del ciclo celular a través de TP53	0,012827469	Formación de colágeno	0,038279854
Cascadas de señalización de la familia MAPK	0,013726982	Metabolismo de los glicosaminoglucanos	0,0385388
Respuestas celulares a estímulos externos	0,01885979	Señalización oncogénica mediada por MAPK	0,039371055
Ciclo del ácido cítrico	0,01885979	Señalización por IGF-1	0,040455837
Señalización mediada por NR1H2 y NR1H3	0,022444743	Cascada PI3K	0,040455837
Cascada de formación del tapón de fibrina	0,022444743	Ciclo de liberación de neurotransmisores	0,047302431
Cadena transportadora de electrones, síntesis de ATP, y producción de calor por proteínas no emparejadas	0,022444743		

El análisis de linfomas foliculares frente a los nódulos reactivos presentó genes con una gran capacidad predictiva fenotípica pero, debido a su heterogeneidad, el número de vías biológicas significativamente alteradas fue bajo. Las 3 con una significación mayor son la

relacionada con la organización de la matriz extracelular ($P=0,005$), la condensación de los cromosomas ($P=0,0054$) y eventos relacionados con ciclinas ($P=0,0076$).

Tabla 3: Listado de las 10 vías biológicas alteradas significativamente en linfoma folicular, ordenadas de forma decreciente según su significación.

VÍA BIOLÓGICA	P VALOR	VÍA BIOLÓGICA	P VALOR
Organización de la matriz extracelular	0,004964	Expresión génica estimulada por YAP-1 y WWTR1 (TAZ)	0,01895
Condensación cromosómica prometáfase	0,005418	Fases G0 y G1 temprana	0,01908
Eventos asociados a ciclinas A/B1 durante la transición G2/M	0,007561	Metabolismo de la biotina (Btn)	0,01971
Transportadores de aminoácidos y oligopéptidos SLC	0,015	Opioide proopiomelanocortina	0,02426
Transporte de cationes/aniones inorgánicos y aminoácidos	0,01518	Señalización mediada por el receptor 5-HT1	0,04921

Finalmente se analizaron los linfomas de Hodgkin frente a los nódulos reactivos. Se obtuvieron 17 vías biológicas alteradas significativamente (Tabla 3). De ellas, las 3 con una mayor significación fueron la señalización por interleucinas ($P=9,51 \cdot 10^{-5}$), la mediada por receptores tirosina quinasa ($P=0,0026$) y señalización por miembros de la familia TGF- β ($P=0,0039$).

Tabla 4: Listado de las 17 vías biológicas alteradas significativamente en linfoma de Hodgkin, ordenadas de forma decreciente según su significación. Destacan las vías con moléculas de carácter inmunológico.

VÍA BIOLÓGICA	P VALOR	VÍA BIOLÓGICA	P VALOR
Señalización por interleucinas	9,51E-05	Señalización mediada por IL-4 y IL-13	0,014917319
Señalización por receptores tirosina quinasa	0,002672787	Regulación de la expresión génica vía NR1H3 & NR1H2 ligada al transporte y flujo de colesterol	0,021623588
Señalización por miembros de la familia TGF- β	0,003956773	Regulación síntesis del colesterol vía SREBP (SREBF)	0,025647158
Sistema neuronal	0,004025832	Red trans-Golgi de formación vesicular	0,026824058
Señalización por el complejo de receptores de TGF- β	0,006306386	Señalización intracelular por segundos mensajeros	0,026824058
Señalización mediada por NR1H2 y NR1H3-	0,007823854	Señalización mediada por familia IL-2	0,044091696
Vía intrínseca para apoptosis	0,007823854	Respuesta celular a stress	0,044879751
Activación plaquetaria, agregación y señalización	0,007823854	Organización de la matriz extracelular	0,044879751
Endocitosis mediada por clatrina	0,014252944		

A continuación, se añade una comparación entre las vías alteradas en linfomas B difusos de célula grande (DLBCL) y linfomas de Hodgkin (Tabla 4).

Tabla 5: Se muestran los *pathways* alterados de manera singular en DLBCL (izquierda) y cHL (derecha). Se excluyeron aquellas vías coincidentes, aunque éstas presentasen diferente significación. En el linfoma de Hodgkin destaca la alteración en la señalización mediada por TGF- β frente a la alteración en la vía RAF/MAPK del DLBCL.

VÍAS BIOLÓGICAS DLBCL	P VALOR	VÍAS BIOLÓGICAS cHL	P VALOR
Señalización vía FLT3	0,0043879	Señalización por miembros de la familia TGF- β	0,00395677
Señalización MAPK1/MAPK3	0,0085605	Señalización por el complejo de receptores de TGF- β	0,00630638
Cascada RAF/MAP quinasa	0,0085605	Vía intrínseca para apoptosis	0,00782385
Desórdenes de transducción de señal	0,0085605	Endocitosis mediada por clatrina	0,01425294
Transmisión a través de sinapsis química	0,0110179	Señalización mediada por IL-4 y IL-13	0,01491731
Transcripción de genes reguladores del ciclo celular a través de TP53	0,0128274	Regulación síntesis del colesterol vía SREBP (SREBF)	0,02564715
Cascadas de señalización de la familia MAPK	0,0137269	Red trans-Golgi de formación vesicular	0,02682405
Ciclo del ácido cítrico	0,0188597	Señalización mediada por familia IL-2	0,04409169
Cascada de formación del tapón de fibrina	0,0224447		
Cadena transportadora de electrones, síntesis de ATP, y producción de calor por proteínas no emparejadas	0,0224447		
Ensamblaje de las cadenas colágenas y otras estructuras multiméricas	0,0289959		
Señalización NCAM para crecimiento de dendritas	0,0327475		
Respuesta celular a stress térmico	0,0331103		
Factores de interacción con VIH	0,0382798		
Señalización AKT activada por PIP3	0,0382798		
Regulación de la respuesta al calor mediada por HSF1	0,0382798		
Eventos relacionados con receptores sensibles a insulina (IRS) mediados por IGF-1	0,0382798		
Formación de colágeno	0,0382798		
Metabolismo de los glicosaminoglucanos	0,0385388		
Señalización por IGFR-1	0,0404558		
Ciclo de liberación de neurotransmisores	0,0473024		

Atendiendo a las vías alteradas significativamente, se orienta el diseño del *knockout* hacia la señalización neuronal, cuya presencia resultó constante en todos los análisis. Además, destaca de otras vías directamente implicadas en el mantenimiento de la homeostasis, metabolismo o el control del ciclo celular (MAPK, PIP₃, fosforilación oxidativa...). Se desestiman estas dianas debido al posible compromiso en la viabilidad

celular secundario a su eliminación. En cambio, la señalización neuronal mediante neurotransmisores aparenta ser singular y más segura de cara a una aplicación clínica. De entre los genes más discriminatorios partícipes de esta vía se elige *HTR3A*, codificante para el receptor de 5-hidroxitriptamina 3 subunidad A (5-HT_{3A}), el cual conforma una proteína de canal pentamérica accionada por ligando (serotonina). Este gen mostró una mayor significación en los linfomas foliculares, con una expresión media casi 3 veces superior (2,95) a las células control del estudio GSE23647, que transformado a log₂ equivale a 1,56 (Figura 4). En los nódulos reactivos, en cambio, la expresión media de *HTR3A* es 1,36 y en forma de log₂ es 0,45. La diferencia entre ambas medias presenta un *P* valor=0.0044 de acuerdo a la T de Student (intervalo de confianza del 95% de que ambas medias son distintas de 0:- 1.8828815; -0.4045498) (Figura 5).

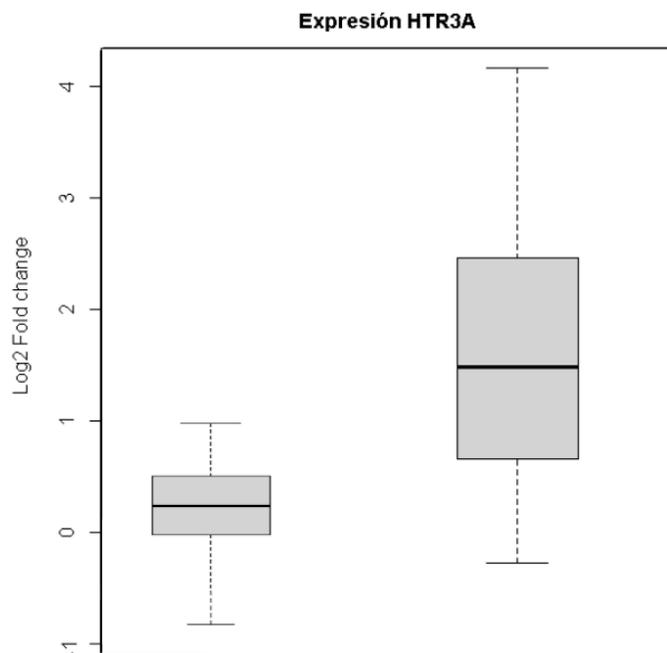


Figura 5: Comparación de la expresión del gen HTR3A en nódulos reactivos (izquierda) y linfomas foliculares (derecha). Este gen se encuentra sobreexpresado con un ratio folicular/reactivo=2,17.

1.2 Diseño del sgRNA para el gen *HTR3A*

Para el presente proyecto se diseñan 8 propuestas de sgRNA cuyo objetivo se encuentra en uno de los exones que componen el gen *HTR3A* humano (Figura 6A). Este gen, localizado en el cromosoma 11q23.2, presenta 3 isoformas, siendo coincidentes en la mayoría de sus segmentos codificantes. No obstante, entre ellas presentan diferencias los exones 1 (no compartido por la tercera isoforma lo que impiden su uso) y 7 de la primera isoforma. El ADN sometido a valoración comprende, por tanto, los exones 2, 3, 4, 5, 6, 7, 8 (comprendido en el 7º exón de la primera isoforma) y 9 (Figura 6B).

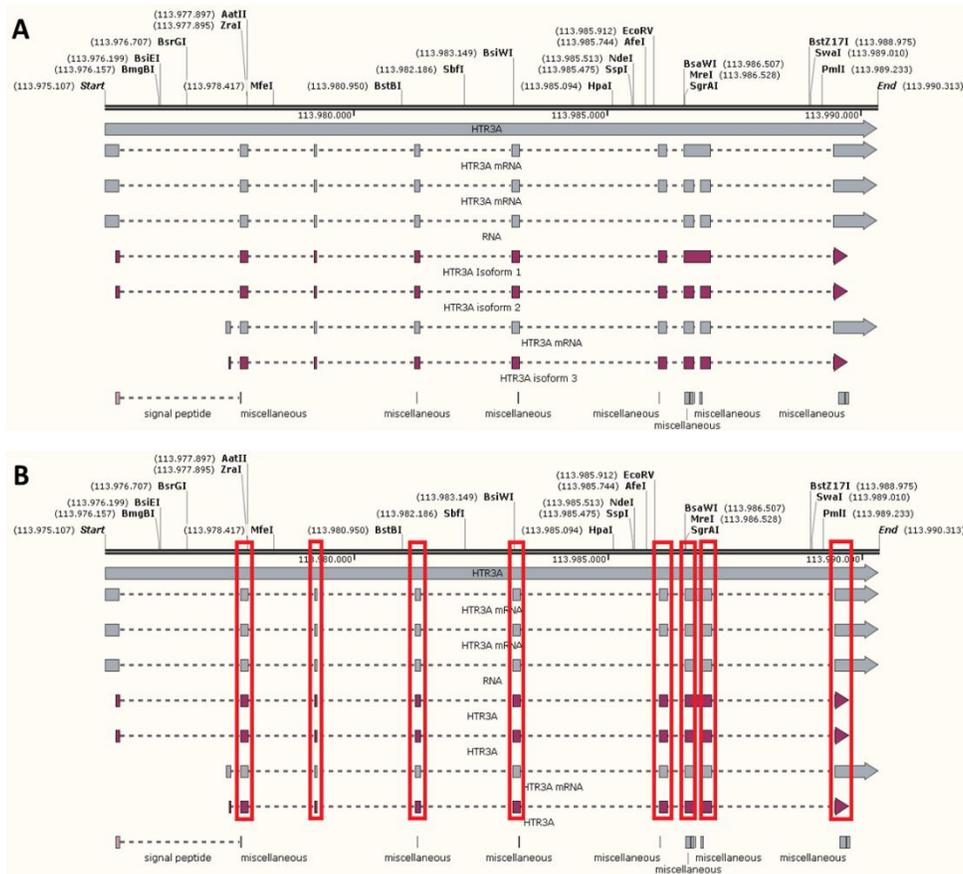


Figura 6: A) Mapa del gen *HTR3A* humano de 15.206 pb. En gris se muestran las regiones implicadas en los ARNm así como otras estructuras nucleicas. En morado se representan las regiones abarcadas por los diferentes exones (8 en la primera isoforma, 9 en las dos restantes). Sobre la línea negra que representa el gen se dispone la localización de los enzimas de restricción. **B) Se destacan en rojo los exones candidatos para el diseño de sgRNAs.** Los exones presentan regiones codificantes compartidas por las tres isoformas del gen *HTR3A*. Imágenes obtenidas y modificadas de *SnapGene*®.

Tras el estudio detenido y la evaluación mediante CRISPOR, se descartan los sgRNAs posibles en los exones 3, 4, 6 y 8, debido a su menor eficacia o la ausencia de sgRNAs sin complementariedad total en la región semilla para 4 o menos *mismatches*. De entre los demás sgRNAs procedentes de los exones 2, 5, 7 y 9 se escogen los siguientes:

➤ **sgRNA 1:** Diana: 5'-AATGACGTCAATGGATACGG-3', reversa. Exón: 2. PAM: TGG.

MIT score: 95. *Doench'16 score:* 73. Deleciones esperadas: Disruptivas del marco de lectura: 55%. Desplazamiento del marco de lectura: 61%. *Off-target events:* 0-0-0-3-38 (para 0, 1, 2, 3 y 4 *mismatches* respectivamente) Ninguno presenta complementariedad completa en la región semilla (*seed region*). Oligonucleótido que sintetizar: 5'-AAUGACGUCAAUGGAUACGG-3' Añadir una guanina en posición +1 en pX330.

➤ **sgRNA 2:** Diana: 5'-GATGCCGAATATACACGTAC-3', reversa. Exón: 5. PAM: GGG.

MIT score: 98. *Doench'16 score:* 65. Deleciones esperadas: Disruptivas del marco de lectura: 63%. Desplazamiento del marco de lectura: 57%. *Off-target events:* 0-0-0-1-18 (para 0, 1, 2, 3 y 4 *mismatches* respectivamente). Ninguno presenta complementariedad completa en la región semilla (*seed region*). Oligonucleótido que sintetizar: 5'-GAUGCCGAAUACACGUAC-3'

➤ **sgRNA 3:** Diana: 5'-GCGTGTGAGAAACGATGATC-3', reversa. Exón: 7. PAM: AGG.

MIT score: 94. *Doench'16 score:* 46. Deleciones esperadas: Disruptivas del marco de lectura: 56%. Desplazamiento del marco de lectura: 62%. *Off-target events:* 0-0-0-1-33 (para 0, 1, 2, 3 y 4 *mismatches* respectivamente). Ninguno presenta complementariedad completa en la región semilla (*seed region*). Oligonucleótido que sintetizar: 5'-GCGUGUCAGAAACGAUGAUC-3'

En el Anexo 1, se muestran imágenes tomadas durante el proceso de diseño del vector plasmídico pX330 para que contenga sgRNA3.

6. DISCUSIÓN

La técnica del estudio masivo de genes usando *microarrays* ha supuesto una revolución en la investigación genómica, permitiendo descubrir no solo genes con expresión aberrante, sino asociaciones funcionales de éstos o comprender su activación en el tiempo. Así, la investigación genética se puede abordar desde diferentes perspectivas empleando el análisis de grandes bases de datos. En primer lugar, se debe tener en cuenta la relación que se establece entre las muestras y los valores obtenidos a partir del estudio genético de éstas. Las muestras pueden ser agrupadas en diferentes fenotipos, véase formas de la enfermedad, respuesta al tratamiento, agresividad... A su vez, los valores obtenidos mediante experimentos *microarray* o RNA-seq permiten investigar el grado de alteración de los genes. Así, para establecer la relación muestras-genes puede plantearse un problema de predicción de fenotipos, como en el análisis del presente estudio. De esta forma, a partir de una serie de muestras dadas, de las cuales su fenotipo es conocido, se analiza qué genes se encuentran desregulados. Una vez determinados los genes con una alteración más significativa, se genera una firma genética capaz de predecir, con la mayor precisión posible, el fenotipo de una muestra nueva. Por tanto, es posible establecer relaciones estadísticas significativas con las que abrir nuevas vías de investigación (26). Además, gracias a la minería de datos estos procedimientos se han podido realizar de forma masiva. Como resultado ha sido posible establecer vías defectivas en múltiples enfermedades, destacando por ejemplo el cáncer de mama triple negativo y el mieloma múltiple (20, 27).

Estudios previos han centrado sus esfuerzos en clasificar los linfomas desde un punto de vista transcriptómico, tanto a nivel global (28) como para determinados linfomas

agresivos (29). Como novedad, en este trabajo se han identificado las principales rutas alteradas en los linfomas más frecuentes. Entre ellas destacan las alteraciones en la señalización intracelular mediada por fosforilación y regulación del ciclo celular, hallazgos compatibles con la proliferación tumoral (30, 31). Entre estas vías observamos el grupo MAPK, PIP₃/AKT, ERK y receptores tirosina quinasa, objeto de diferentes estudios en linfomas (32, 33). En relación con la proliferación celular, se encuentran las vías de activación y señalización plaquetaria. Así, los factores secretados por las plaquetas (PDGF, PF4, VEGF), como ellas mismas se han relacionado con la progresión tumoral en numerosos estudios (34, 35). La señalización mediada por interleucinas ha sido una de las vías más desreguladas. En el caso de los linfomas de Hodgkin aparecen implicadas las interleucinas 2, 4 y 13, así como varios genes que codifican moléculas de señalización características del sistema inmune, tales como las quimiocinas. Concretamente en el linfoma de Hodgkin, estas moléculas quimiotácticas presentan una desregulación genética altamente significativa. Una de ellas es CCL17 cuya expresión es 12 veces mayor que en las células hematopoyéticas control y 8,3 veces mayor que los nódulos reactivos. Publicaciones previas respaldan estos hallazgos (36, 37), poniendo de manifiesto la interacción entre las células Reed-Sternberg y el estroma tumoral, así como la modulación que ejercen sobre la respuesta inmune. Su relevancia se ha ratificado mediante la inhibición de su receptor CCR4, disminuyendo la proliferación tumoral(38).

El sistema CRISPR/Cas presenta numerosas aplicaciones en la actualidad. Es una forma sencilla de generar *knockouts*, especialmente eficaz en ensayos *in vitro*. Además, permite corregir genes defectuosos mediante la recombinación homóloga, así como la

modificación de mutaciones puntuales mediante endonucleasas Cas modificadas. El potencial que presenta hace del CRISPR/Cas una de las herramientas más prometedoras en la terapéutica futura, encontrándose en fase de ensayo clínico para el tratamiento de la beta talasemia o la anemia de células falciformes (39). En este estudio se ha seleccionado un gen de la vía de señalización a través de neurotransmisores para realizar el diseño de sgRNA. Concretamente en la serotonina, el gen codificante del receptor 3 subunidad A (*HTR3A*) que se encuentra significativamente sobreexpresado en linfomas foliculares. No es la primera vez que se analiza el papel de la serotonina y sus receptores en el desarrollo tumoral. Así, la expresión de *HTR1A* ha sido puesta de manifiesto en DLBCL o del manto, y el empleo de antagonistas de este receptor han mostrado capacidad para inhibir el crecimiento tumoral (40). Por su parte, la desregulación de *HTR3A* ha sido observada en linfomas derivados del centro germinal (DLBCL, FL), correlacionándose su expresión con el grado tumoral en los linfomas foliculares (41). Además, la sobreexpresión de *HTR3A* ha mostrado ser marcador de agresividad en adenocarcinomas de pulmón, activando la vía ERK; así como su inhibición capaz de disminuir la proliferación tumoral (42). Efectos similares han sido obtenidos en adenocarcinomas colorrectales (43). Sin embargo, no hemos encontrado trabajos previos sobre el efecto que supone el *knockout* de *HTR3A* en linfomas que lo sobreexpresen. Finalmente, la investigación orientada a clarificar su papel en la tumorigénesis de estos linfomas, mediante un experimento que emplease modelos CRISPR/Cas como los propuesto en el presente estudio, podría suponer un avance en la determinación de *HTR3A* como factor pronóstico y el diseño de nuevos abordajes terapéuticos.

7. CONCLUSIONES

- 1- Los resultados obtenidos mediante el análisis estadístico basado en la predicción de fenotipos presentan coherencia científica y coinciden con resultados previos obtenidos por otros grupos a nivel internacional. Así, el método se muestra válido para explorar la desregulación génica tumoral de forma rápida y automatizada.
- 2- Los linfomas presentan una desregulación amplia y heterogénea, que afecta a vías de señalización básicas para la proliferación celular como MAPK, PIP₃/AKT, ERK y receptores tirosina quinasa. No obstante, también los mensajeros extracelulares (p.ej. quimiocinas, interleucinas) se ven alterados, reafirmando la importancia del estroma para el desarrollo tumoral.
- 3- Se han encontrado vías biológicas alteradas capaces de abrir nuevas líneas de investigación que aspiren a inhibir el desarrollo tumoral. Además, algunas de estas vías cuentan con fármacos que las reconocen específicamente como diana (FLT3, receptor 5-HT₃, entre otras). Por tanto, este trabajo sienta las bases estadísticas para justificar nuevos proyectos pre-clínicos.
- 4- La vías relacionadas con la señalización neuronal se muestra significativamente desreguladas en los linfomas. Entre los receptores activados por serotonina, el gen *HTR3A* se encuentra sobreexpresado en linfomas foliculares (Log₂ FC = 1,56, P=0,0044). Tomando como referencia los resultados positivos obtenidos por estudios previos, la supresión *in vitro* de este gen habría de ser explorada para clarificar su potencial terapéutico.

8. BIBLIOGRAFÍA

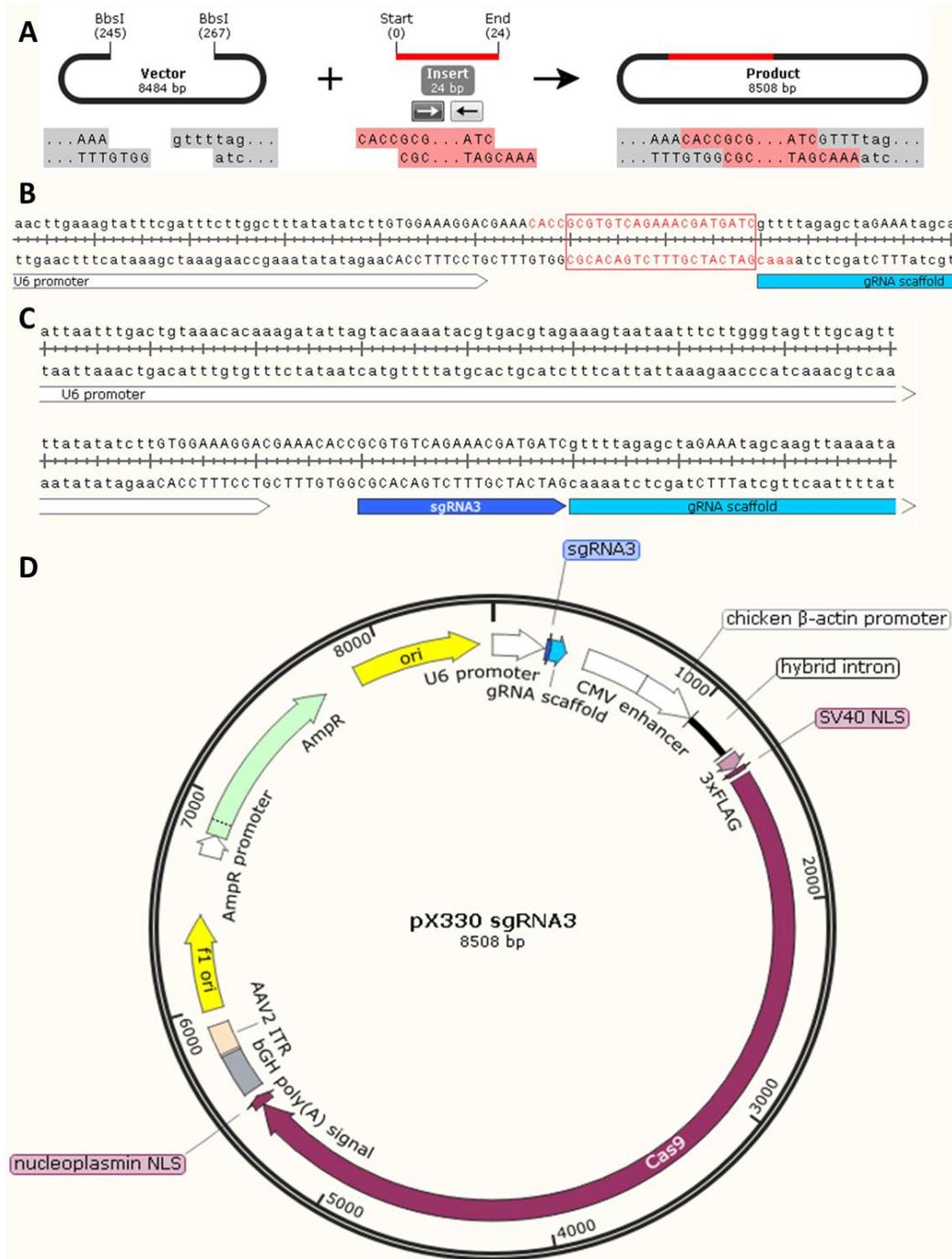
1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015;136(5):E359-86.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
3. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375-90.
4. Jiang M, Bennani NN, Feldman AL. Lymphoma classification update: T-cell lymphomas, Hodgkin lymphomas, and histiocytic/dendritic cell neoplasms. *Expert review of hematology*. 2017;10(3):239-49.
5. Boyiadzis M, Bishop MR, Abonour R, Anderson KC, Ansell SM, Avigan D, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of hematologic malignancies: multiple myeloma, lymphoma, and acute leukemia. *Journal for immunotherapy of cancer*. 2016;4:90.
6. Klee EW. Data mining for biomarker development: a review of tissue specificity analysis. *Clinics in laboratory medicine*. 2008;28(1):127-43.
7. Raja K, Patrick M, Gao Y, Madu D, Yang Y, Tsoi LC. A Review of Recent Advancement in Integrating Omics Data with Literature Mining towards Biomedical Discoveries. *International journal of genomics*. 2017;2017:6213474.
8. Branco I, Choupina A. Bioinformatics: new tools and applications in life science and personalized medicine. *Applied microbiology and biotechnology*. 2021;105(3):937-51.
9. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature genetics*. 2001;29(4):365-71.
10. Wiltgen M, Tilz GP. DNA microarray analysis: principles and clinical impact. *Hematology*. 2007;12(4):271-87.
11. Tang PW, Chua PS, Chong SK, Mohamad MS, Choon YW, Deris S, et al. A Review of Gene Knockout Strategies for Microbial Cells. *Recent Pat Biotechnol*. 2015;9(3):176-97.

12. Chemistry Nobel Honors CRISPR, an «Essential» Tool for Cancer. *Cancer Discov.* 2020; 10(12):1782-1783.
13. Mojica FJM, Díez-Villaseñor C. The on-off switch of CRISPR immunity against phages in *Escherichia coli*. *Molecular Microbiology*. septiembre de 2010; 77(6):1341-5.
14. Altenbuchner J. Editing of the *Bacillus subtilis* Genome by the CRISPR-Cas9 System. *Applied and Environmental Microbiology*. 2016; 82(17):5421-7.
15. Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. *Quant Biol.* 2014; 2(2):59-70.
16. Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*. 2018; 361(6408):1259-62.
17. Ding W, Zhang Y, Shi S. Development and Application of CRISPR/Cas in Microbial Biotechnology. *Frontiers in bioengineering and biotechnology*. 2020; 8:711.
18. Dudoit S, Yang JYH. Bioconductor R Packages for Exploratory Analysis and Normalization of cDNA Microarray Data. En: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL, editores. *The Analysis of Gene Expression Data: Methods and Software* [Internet]. New York, NY: Springer; 2003 [citado 11 de octubre de 2020]. p. 73-101. (Statistics for Biology and Health).
19. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst.* 2016;12(2):477-9.
20. Cernea A, Fernandez-Martinez JL, de Andres-Galiana EJ, Fernandez-Ovies FJ, Alvarez-Machancoses O, Fernandez-Muniz Z, et al. Robust pathway sampling in phenotype prediction. Application to triple negative breast cancer. *BMC bioinformatics*. 2020; 21(2):89.
21. Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome biology*. 2016;17(1):148.
22. Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology*. 2015; 33(9):985-9.
23. Fujihara Y, Ikawa M. CRISPR/Cas9-based genome editing in mice by single plasmid injection. *Methods in enzymology*. 2014;546:319-36.

24. Iijima O, Fukano H, Takahashi H, Shirai M, Suzuki Y. A purine at +2 rather than +1 adjacent to the human U6 promoter is required to prepare effective short hairpin RNAs. *Biochemical and biophysical research communications*. 2006; 350(4):809-17.
25. Sakai R, Esaki Y, Hasuwa H, Ikawa M, Lo P, Matsuura R, et al. Knockout of Cytidine Monophospho-N-Acetylneuraminic Acid (CMP-NeuAc) Hydroxylase From Porcine Endothelial Cells by a CRISPR System. *Transplantation proceedings*. 2016; 48(4):1320-2.
26. Cernea Corbeanu A, Fernández Martínez JL, Andrés Galiana EJ, Fernández Ovies FJ, Fernández Muñiz MZ, Álvarez Machancoses Ó, et al. Comparison of Different Sampling Algorithms for Phenotype Prediction. *Bioinformatics and Biomedical Engineering Lecture Notes in Computer Science*, Springer. 2018; 10814:33-45.
27. Fernandez-Martinez JL, de Andres-Galiana EJ, Fernandez-Ovies FJ, Cernea A, Kloczkowski A. Robust Sampling of Defective Pathways in Multiple Myeloma. *International journal of molecular sciences*. 2019;20(19).
28. Loi TH, Campain A, Bryant A, Molloy TJ, Lutherborrow M, Turner J, et al. Discriminating lymphomas and reactive lymphadenopathy in lymph node biopsies by gene expression profiling. *BMC Med Genomics*. 2011; 4:27.
29. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000, 403: 503-511.
30. Masliah-Planchon J, Garinet S, Pasmant E. RAS-MAPK pathway epigenetic activation in cancer: miRNAs in action. *Oncotarget*. 2016;7(25):38892-907.
31. Singh V, Ram M, Kumar R, Prasad R, Roy BK, Singh KK. Phosphorylation: Implications in Cancer. *The protein journal*. 2017;36(1):1-6.
32. Schmidt J, Ramis-Zaldivar JE, Nadeu F, Gonzalez-Farre B, Navarro A, Egan C, et al. Mutations of MAP2K1 are frequent in pediatric-type follicular lymphoma and result in ERK pathway activation. *Blood*. 2017;130(3):323-7.
33. Locatelli SL, Careddu G, Stirparo GG, Castagna L, Santoro A, Carlo-Stella C. Dual PI3K/ERK inhibition induces necroptotic cell death of Hodgkin Lymphoma cells through IER3 downregulation. *Scientific reports*. 2016; 6(1):35745.
34. Wang S, Li Z, Xu R. Human Cancer and Platelet Interaction, a Potential Therapeutic Target. *Int J Mol Sci*. 2018; 19(4):1246.
35. Peterson JE, Zurakowski D, Italiano JE, Michel LV, Connors S, Oenick M, et al. VEGF, PF4 and PDGF are elevated in platelets of colorectal cancer patients. *Angiogenesis*. 2012; 15(2):265-73.

36. Hnatkova M, Mocikova H, Trneny M, Zivny J. The biological environment of Hodgkin's lymphoma and the role of the chemokine CCL17/TARC. Prague medical report. 2009;110(1):35-41.
37. Niens M, Visser L, Nolte IM, van der Steege G, Diepstra A, Cordano P, et al. Serum chemokine levels in Hodgkin lymphoma patients: highly increased levels of CCL17 and CCL22. British journal of haematology. 2008;140(5):527-36.
38. Ishida T, Ishii T, Inagaki A, Yano H, Kusumoto S, Ri M, et al. The CCR4 as a novel-specific molecular target for immunotherapy in Hodgkin lymphoma. Leukemia. 2006;20(12):2162-8.
39. Frangoul H, Altshuler D, Cappellini MD, Chen Y-S, Domm J, Eustace BK, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. N Engl J Med. 2021; 384(3):252-60.
40. Kolan SS, Lidstrom T, Mediavilla T, Dernstedt A, Degerman S, Hultdin M, et al. Growth-inhibition of cell lines derived from B cell lymphomas through antagonism of serotonin receptor signaling. Scientific reports. 2019; 9(1):4276.
41. Rinaldi A, Chiaravalli AM, Mian M, Zucca E, Tibiletti MG, Capella C, et al. Serotonin receptor 3A expression in normal and neoplastic B cells. Pathobiology : journal of immunopathology, molecular and cellular biology. 2010;77(3):129-35.
42. Tone M, Tahara S, Nojima S, Motooka D, Okuzaki D, Morii E. HTR3A is correlated with unfavorable histology and promotes proliferation through ERK phosphorylation in lung adenocarcinoma. Cancer science. 2020;111(10):3953-61.
43. Tang J, Wang Z, Liu J, Zhou C, Chen J. Downregulation of 5-hydroxytryptamine receptor 3A expression exerts an anticancer activity against cell growth in colorectal carcinoma cells in vitro. Oncology letters. 2018;16(5):6100-8.

9. ANEXO 1



Anexo 1. Representación del proceso de introducción del sgRNA3 en pX330. A) Se digiere pX330 con el enzima de restricción BbsI, uniéndose mediante complementariedad de bases a los extremos cohesivos del sgRNA3. B) Se observa el resultado de ligar ambos fragmentos de ADN, resaltándose en el recuadro las 20 bases que componen el sgRNA3 y su hebra complementaria. C) Muestra del resultado final del procedimiento, con el sgRNA introducido correctamente. D) Vista general del plásmido finalizado el proceso, encontrándose listo para su síntesis.