Mejora de la detección de neoantígenos tumorales mediante el uso de tecnología de aprendizaje automático

Jokin Eguía Sánchez

Trabajo Fin de Máster. Curso 2019-2020

Director: Esteban Veiga Chacón
Codirector: Carlos Oscar Sánchez Sorzano
Tutor: Daniel Aguirre de Carcer García
Lugar de realización: Centro Nacional de Biotecnología (CNB-CSIC)

Índice

1. Abstract

According to the WHO (World Health Organization), cancer is the second leading cause for death in the world. In 2012 around 14.1 million new cases were detected, and 8.2 million deaths were recorded. It is estimated that by 2025, 20 million cases will be detected anually (Ferlay *et al.,*2015). Cancer is a disease caused by a genetic disorder, be it hereditary, pathogen-driven such as by viruses, environment-driven, etc., which can affect any cell in the human body. Current cancer research has been full of steps in the right direction, although some have also been deceitful. Apart from classic cancer treatments such as chemotherapy and radiotherapy, a new revolutionary field has arisen in oncology, based on the use of immunotherapy, specifically for poor-prognosis tumors. To get an idea of the impact of this new field, *Science Journal* chose immunotherapy as 'achievement of the year' back in 2013. One of the key elements the immune system uses in order to be able to destroy malignant cells are neoantigens, antigens present only in the MHC-1 of tumor cells, which are not present in healthy cells, since their origin is due to mutations during the process of celular malignancy (or the expression of phoetal antigens). This allows the immune system to distinguish between healthy and tumor cells. Developing accurate techniques that detect neoantigens is fundamental in order to achieve efficient personalized therapies. One such way of finding these therapies, which has not been much explored thus far, would be the development of self-learning neural networks, which, once trained, would be able to detect neoantigenic fragments in RNA sequences from tumor samples.

2. Resumen

Según la OMS el cáncer es la segunda causa de muerte en el mundo. En 2012 se detectaron al rededor de 14,1 millones de casos nuevos de cáncer y 8,2 millones de defunciones, y se estima que para el 2025 se detectarán 20 millones de nuevos casos anualmente (Ferlay *et al.,* 2015). El cáncer es una enfermedad que es desencadenada por algún desorden genético, pudiendo ser este hereditario, provocado por patógenos como virus, por factores ambientales etc. que puede afectar a cualquier célula del cuerpo humano. El devenir de las investigaciones oncológicas ha estado repleto de pequeños avances, algunos de ellos nugatorios. Además de los métodos de tratamientos clásicos

que existen para tratar al cáncer, como la radioterapia o quimioterapia, está surgiendo un nuevo campo que esta revolucionando el campo de la oncología que consiste en el uso de inmunoterapias que permiten el tratamiento de los tumores con un pronóstico muy malo. Para hacerse una idea del impacto de este nuevo campo *Science Journal* eligió la inmunoterapia como 'logro del año' en el 2013. Uno de los elementos clave que utiliza el sistema inmune para poder destruir las células malignas es la aparición de neoantígenos, que son antígenos presentados en el MHC I de las células tumorales que no se encuentran en una célula sana, ya que su origen se debe a las mutaciones somáticas que van apareciendo en el proceso de malignización celular, o a la expresión de antígenos fetales. Esto permite al sistema inmune discernir lo propio (células sanas) de lo extraño (antígenos tumorales). Desarrollar sistemas eficaces para detectar estos neoantígenos es vital para poder lograr unas terapias personalizadas eficientes. Para lograr esta detección una posible opción, que aún no ha sido muy explotada, sería el desarrollo de redes neuronales de aprendizaje automatizado. Las cuales, una vez entrenadas serían capaces de detectar posibles neoantígenos a partir de la secuenciación de ARN de una muestra tumoral.

3. Objetivos

El principal objetivo de este trabajo consiste en la exploración preliminar del desarrollo de una estrategia de vacunación neoantigénica para tumores. Para ello el proyecto se divide en dos objetivos.

Es decir, se establecerán las bases para generar una plataforma que permita identificar neoantígenos en tumores clínicos (humanos) a partir de la secuencia de RNA del tejido tumoral. Este trabajo no parte desde 0, ya que se buscarán nuevos modelos y arquitecturas que puedan mejorar o complementar un trabajo previo realizado en el laboratorio de acogida donde se generaron algoritmos para predecir neoantígenos usando un modelo de melanoma agresivo de ratón, en este caso aplicado al sistema MHC humano. Hemos empezado, desarrollando varias bases de datos con fragmentos de proteínas y donde se discierne si son o no inmunogénicos, para después entrenar las redes neuronales diseñadas para la identificación de fragmentos inmunogénicos que podrían ser candidatos a neoantígenos. Para ello se desarrollarán dos tipos de redes

neuronales diferentes, una de tipo convolucional y otra de tipo red neuronal recurrente, concretamente una LSTM (Long Short Term Memory). La primera de ellas, la convolucional originalmente, surgió para la identificación de imágenes. Consiste en un tipo de red neuronal profunda que presenta diferentes capas. Su funcionamiento original consistía en que cada capa detectaba patrones de unas características de las imágenes, cuanto más profunda la capa, características más complejas. Utilizando un modelo similar las convolucionales unidimensionales, utilizadas para analizar secuencias de letras, por ejemplo textos, se puede lograr algo parecido, cuando más profundas sean las capas, distinguirá características más complejas de la información que se introduzca. Por ejemplo es utilizada para la traducción o para análisis de sentimientos. No posee memoria como tal, no obstante dependiendo de como se estructure puede almacenar información previa, donde los datos de secuencias se modelan como series temporales, en las que cada elemento está precedido de otro y seguido de otro. Pese que el usó de este modelo de red neuronal no es común para la detección de neoantígenos, las características comentadas avalan su utilidad para enfrentar este tipo de problemas. El otro modelo de red la LSTM (Long Short Term Memory) ha sido ya utilizada para la identificación de neoantígenos (Shao *et al.,* 2020). Esta arquitectura, a diferencia de la anterior, sí que puede almacenar información sobre los elementos precedentes, lo cual la convierte en una candidata idónea para detectar patrones a lo largo de una secuencia. Un uso habitual que recibe es la de ser usada para traducir. Es evidente entonces porque el uso de este tipo de redes neuronales puede ser útil para este tipo de problemas, el desarrollo de ambas puede ser útil para constatar las diferencias que existen entre la predicción de ambas redes y ver su eficacia e incluso en un futuro combinar ambas para dar lugar a redes más complejas y con mayor capacidad de predicción.

Debido a la naturaleza integrativa de este trabajo, se van a combinar técnicas *in silico* con experimentales. De esa forma, la segunda parte de los objetivos consistiría en testar los antígenos que predijo una red neuronal desarrollada el año anterior en el laboratorio de acogida para la detección de neoantígenos murinos. En este caso concreto para la detección de neoantígenos de B16, un modelo de melanoma agresivo de ratón. Se parte de las predicciones realizadas por esta red neuronal y se constataría su eficacia en vivo mediante vacunación y así poder comprobar la funcionalidad, en un experimento

de prueba de concepto, de las redes neuronales para la detección de neoantígenos.

4. Introducción

El sistema inmune es capaz destruir agentes extraños que han logrado entrar y trasvasar las barreras naturales que protegen a un organismo. Clásicamente se divide en dos, inmunidad innata e inmunidad adaptativa. Dentro de la inmunidad innata destacan las células fagocíticas y los mecanismos que estas desencadenan encargadas de defender al organismo de forma más inespecífica que la inmunidad adaptativa. Las células de la inmunidad innata sin embargo si que son capaces de reconocer moléculas conservadas en microorganismos; reconocen lo que se conoce como patrones moleculares asociados a patógenos (PAMPs en sus siglas en inglés) que son esenciales para la supervivencia y la patogenicidad (Kumar *et al.*,, 2011). Los PAMPs se detectan mediante sensores del huésped conservados evolutivamente, conocidos como receptores de reconocimiento de patógenos (PRR) (Kumar *et al.*,, 2011). Aunque está descrito que este tipo de inmunidad no desarrolla memoria, en los últimos años se ha descubierto que las células de la inmunidad innata también pueden generar una especie de memoria llamada inmunidad entrenada, donde en una segunda exposición de patógenos (no necesariamente el mismo patógeno) responden de forma más eficaz, se ha visto que esta memoria está orquestada por la reprogramación epigenética (Netea *et al.,* 2016). A este sistema innato pertenecen los macrófagos, los neutrófilos y las células dendríticas entre otros. Por otro lado, el sistema de inmunidad adaptativa, a diferencia del anterior, es específico. Las células de este tipo de inmunidad son capaces de identificar elementos extraños (no propios), llamados antígenos, y eliminan al ente biológico que los porta. La inmunidad adaptativa genera memoria inmunitaria, que protege al organismo en futuros encuentros con el mismo antígeno. El sistema inmune responderá de manera mucho más rápida y eficaz en estos encuentros posteriores con los antígenos que han desencadenado la memoria. Las células que constituyen este sistema son los linfocitos T y B. La labor de el sistema inmune en general, no es sólo la de proteger frente a agentes extraños, también protege frente a células que han sido infectadas por virus o bacterias, induciendo la muerte de estas mediante las del sistema inmune innato, células NK (Natural Killer) o mediante la activación de linfocitos T que destruyen la

célula infectada.

El sistema inmune no sólo es capaz de detectar patógenos y eliminar las células infectadas, sino que tanto el sistema inmune innato como la inmunidad adaptativa, han demostrado tener la capacidad para detectar, combatir e incluso eliminar el cáncer. Las células T CD8+ son las mas importantes en la lucha antitumoral porque son las que principalmente se encargan de destruir los tumores. Las células dendríticas capturan antígenos de los tumores, y se los presentan a las células T CD8+ vírgenes en los ganglios linfáticos, que se activan. Este mecanismo conocido como presentación cruzada es fundamental para la destrucción del tumor. Los infocitos T CD8+ son capaces de viajar hasta el tumor, reconocer los antígenos tumorales (neoantígenos) presentados en los MHC-I, y destruir estas células (). En resumen, el sistema inmune adaptativo puede detectar a las células tumorales, como consecuencia del proceso de malignificación que sufren estas, que conlleva a la acumulación de mutaciones y consecuentemente la presentación de antígenos que sólo están presentes en los tumores y no en los tejidos sanos. También puede suceder que en las células tumorales se expresen antígenos fetales, que no deberían estar en un individuo adulto dando lugar a la misma respuesta.

Las células dendríticas o los macrófagos entre otros, suelen forma parte de la respuesta inmune innata antitumoral (Weinberg, 2014). Aunque también pueden jugar el papel contrario y forma parte de microambiente tumoral que las utiliza para apagar la respuesta mediada por los linfocitos T CD8 (DeNardo, & Ruffell, 2019).

Es bien conocido que según el tipo de microambiente el cáncer puede tener buena o mala prognosis. Por ejemplo, si el microambiente está enriquecido en macrófagos M2 y linfocitos Tregs la prognosis es mala (Ino *et al.,* 2013), sin embargo un infiltrado rico en linfocitos T CD8+ y linfocitos T CD4+ th1 suele estar relacionado con una buena prognosis (Anz *et al.,* 2011).

Igualmente, si los infiltrados tumorales son ricos en células B, esto suele estar asociado a buena prognosis (Wouters& Nelson, 2018, Helmink *et al.,* 2020). No está muy claro el motivo, pero parece que está relacionado con la capacidad de presentación de antígenos de estas células B.

También es conocido desde hace tiempo que la células NK estudian las superficies de las células autólogas con el fin de detectar una expresión aberrante de las moléculas MHC I y/o marcadores de estrés celular que pueda indicar un posible foco tumoral. Si la célula NK descubre una célula tumoral le inducirá muerte celular o provocará una respuesta inmune contra esta (Waldhauer & Steinle, 2008; Delves *et al.,* 2008).

Como se acaba de comentar la capacidad de ambos sistemas inmunes para identificar no solo agentes extraños, sino también a células tumorales, ha dado lugar a una nueva rama para el tratamiento del cáncer, conocido como inmunoterapia. El cáncer es una enfermedad que puede originarse a partir de cualquier tipo celular, y que tienen en común, entre otras características, la proliferación descontrolada de la población celular y la capacidad última de migrar y originar focos tumorales secundarios, dando lugar a la metástasis, evasión del sistema inmune... (Hanahan & Weinberg, 2011). En la siguiente figura 1 realizada por (Hanahan & Weinberg, 2011). se muestran los sellos distintivos que suelen caracterizar al cáncer.

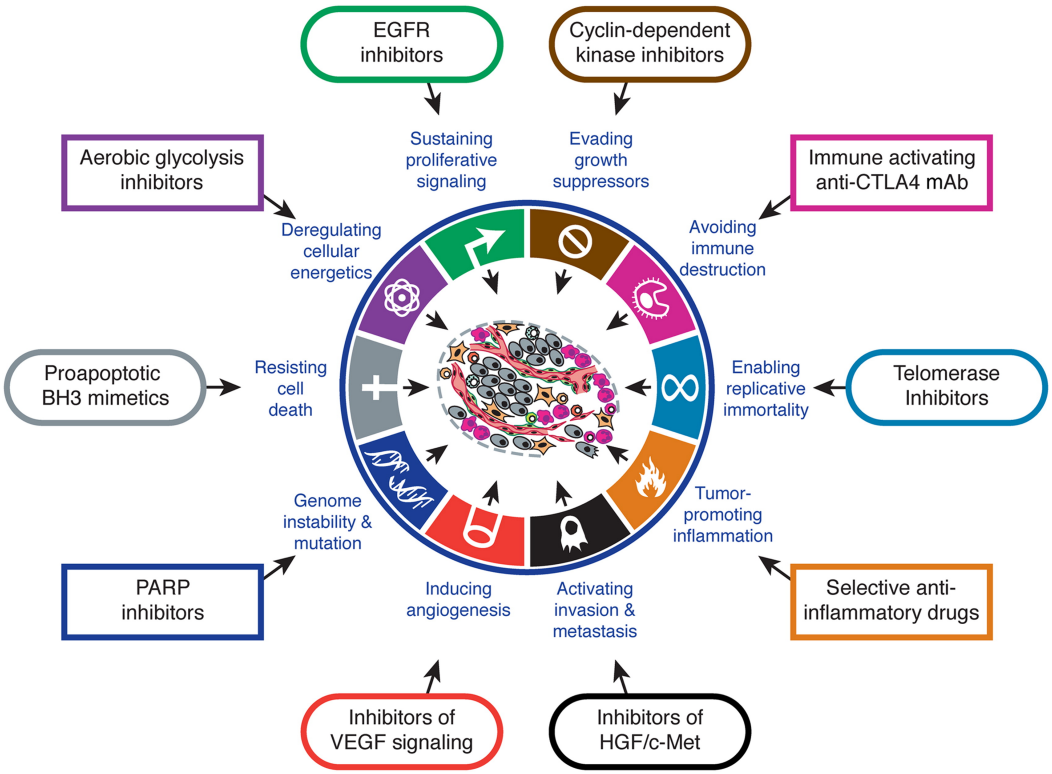

Figura 1: Características del cáncer. Fuente: Hanahan & Weinberg, 2011

Cualquier célula puede dar lugar a la formación de un cáncer, y este puede originarse por una serie de mutaciones que es distinta, no sólo en cada tipo de cáncer, sino

también dentro de la misma población del tumor. Esto supone, un verdadero reto y un paradigma para los investigadores, ya que este heterogéneo origen de la enfermedad dificulta mucho la búsqueda de un tratamiento común. Es por esto que en los últimos años a surgido el concepto de, terapia personalizada, que consiste en aplicar tratamientos terapéuticos en función de las características genómicas y moleculares del tumor de cada paciente (SEOM: Sociedad Española de Oncología Médica, 2019). Es decir, se diseña un tratamiento específico para el paciente. Evitando así el uso de terapias generalistas que dependiendo del tipo de cáncer podrían resultar ineficaces. El tema objeto de este trabajo consiste precisamente en un tratamiento personalizado, en una forma de la inmunoterapia.

La importancia que presenta el sistema inmune adaptativo detectando y combatiendo ciertos tipos de cáncer, está más que corroborada, especialmente para los linfocitos T (Schumacher & Schreiber, 2015). Se ha constatado que ratones con un sistema inmune deprimido son más susceptibles al surgimiento de cánceres espontáneos inducidos por carcinógenos si se comparan con ratones sanos inmunocompetentes (Schreiber *et al.,* 2011). Las células del sistema inmune adaptativo son capaces de reconocer las células tumorales malignas gracias a la presentación antigénica en el constexto de MHC (complejo mayor de histocompatibilidad/Histocompatibility Mayor Complex). Existen dos tipos de MHC, los de clase I MHC I y clase II MHC II. En este trabajo se trata con los de clase I. Los genes de MHC codifican para una glicoproteína que estará constituida de dos polipéptidos, uno, conocido como alfa tiene tres dominios y el otro es la beta microglobulina. El MHC I sirve de soporte para la presentación, en la superficie de todas las células nucleadas del cuerpo, de fragmentos de proteínas propias, que serán presentadas a los linfocitos T, y servirá para que estos puedan reconocer lo propio de lo extraño.

Un linfocito T puede reconocer una célula tumoral si esta presenta vía MHC-I fragmentos peptídicos que no deberían estar presentes en el organismo, por ejemplo, de origen fetal (Schumacher & Schreiber, 2015) o antígenos que se generan como consecuencia de una las mutaciones somáticas generadas durante el proceso de conversión en una célula maligna. Estos neoantígenos, pueden desencadenar el rechazo del tejido tumoral (Schumacher & Schreiber, 2015; Desrichard *et al.,* 2016). También se consi-

deran neoantígenos a los epítopos derivados de marcos de lecturas virales abiertas en cánceres de origen virales.

La frecuencia con la que surgen neoantígenos depende del tipo de cáncer y de la cantidad de mutaciones que sufra este. En la fig. 2 se puede apreciar la frecuencia con la que suelen surgir neoantígenos en distintos tipos de cánceres (Schumacher & Schreiber, 2015).

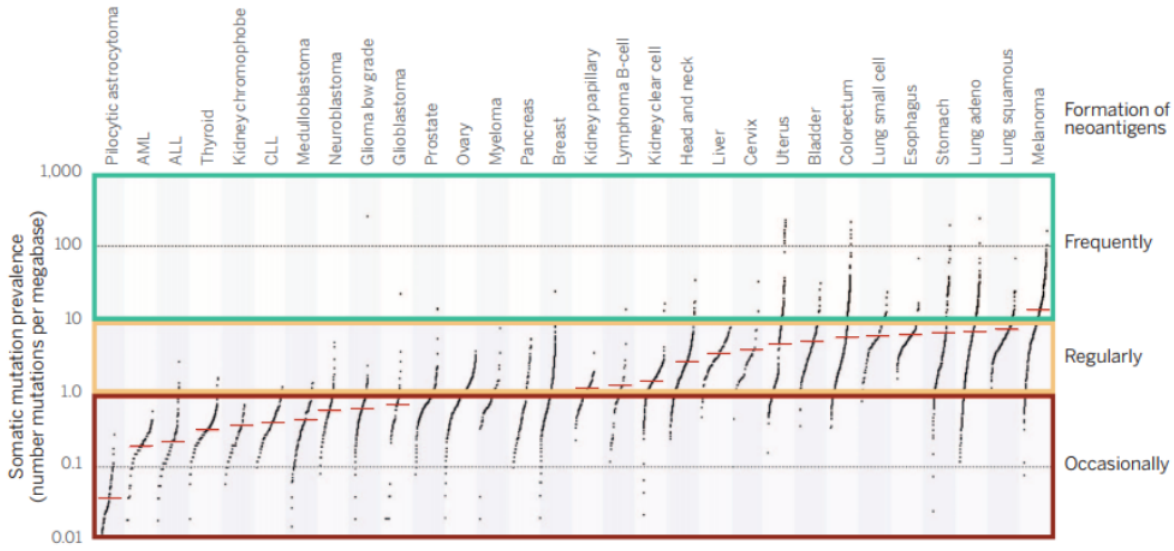

Figura 2: Características del cáncer. Fuente: Schumacher & Schreiber, 2015

Uno de los cánceres que más respuesta inmune suele provocar y que genera más neoantígenos es el melanoma, que es el cáncer en el que más se ha estudiado la respuesta inmune anticancerígena (Schumacher & Schreiber, 2015). No obstante, desafortunadamente, no es suficiente con que surjan estas moléculas en el tumor, las linfocitos T deben de ser capaces de reconocerlo y responder adecuadamente.Durante el proceso de activación de los linfocitos, al mismo tiempo, expresa tiempo una serie de proteínas que tienen la función de "apagarlo", para evitar una inflamación exagerada. Algunos tumores aprovechan esta características de los linfocitos inactivándolos y que así no puedan realizar correctamente su función. Se sabe, por ejemplo que algunos tumores que son reconocidos por las células T, sobreexpresan el ligando inhibitorio PD-l1 evitando así que las células T, que expresan PD-1 que se une a PD-l1, acaben con ellas (Caldwell et al., 2017) fig. 3.Así mismo, le microambiente tumoral suele ejercer una pa-

pel inmunosupresor impidiendo el correcto funcionamiento de los linfocitos (Farhood *et al.,* 2019).

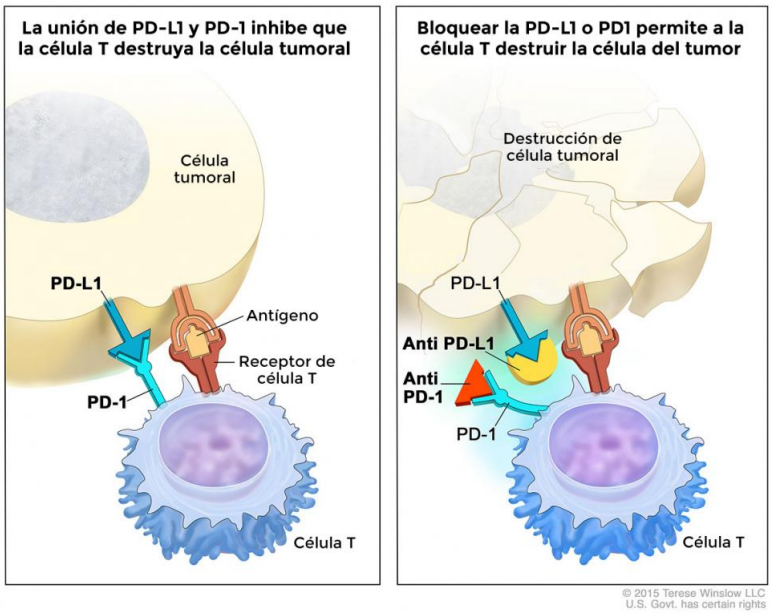

Figura 3: Características del cáncer. Fuente: NIH: National Cancer Institute, 2020.

Existen varias formas para lograr que el sistema inmune reconozca los neoantígenos tumorales y provoque rechazo de cáncer. Una de ellas es la inhibición de puntos control inmunitario, que no es más que evitar que el linfocito se inactive por la expresión de algún ligando y pueda así seguir realizando su función si ha logrado reconocer el tumor, como el que se ha comentado anteriormente (Caldwell et al., 2017; Desrichard et al., 2016; NIH: National Cancer Institute, 2020). De hecho, existen varias terapias basadas en evitar que los tumores activen estas moléculas inhibitorias en los linfocitos, usando anticuerpos que impiden la union de los ligandos con sus receptores (anti PD1, anti ligando de PD1, anti CTLA4, anti LAG3 etc). Otra opción puede ser la expansión *Ex vivo* de linfocitos T que reconozcan debidamente neoantígenos de los tumores (Schumacher & Schreiber, 2015).

Un nuevo método, aun en fase experimental que ha puesto a punto el laboratorio de acogida, ha demostrado ser eficaz en tratamiento de tumores solidos. Este método se basa en el uso de linfocitos T CD4+ "entrenados" con bacterias que expresan antígenos tumorales (Cruz-Adalia *et al.,* 2017). El laboratorio de acogida ha demos-

trado que los linfocitos T CD4+ tienen la capacidad de capturar bacterias desde una célula dendrítica (DC) infectada (Cruz-Adalia *et al.,* 2014), hecho que rompe el dogma sobre la separación de los roles entre la inmunidad innata y adaptativa. Nótese que los linfocitos T CD4+ son el paradigma de la inmunidad adaptativa y se demostró que eran células fagocíticas que capturaban bacterias por transfagocitosis. Es más, eliminaban de forma muy eficaz las bacterias que capturaban y presentaban de forma cruzada antígenos bacterianos a linfocitos T CD8+ vírgenes, activándolos mejor que las propias DC. Estas nuevas habilidades de los linfocitos T CD4+ "entrenados" por bacterias se suponían que eran exclusivas de las células de la inmunidad innata. Además, los linfocitos T CD8+ activados por los linfocitos T CD4+ entrenados como APC, generaban memoria central (mucho más resistentes al "cansancio" inducido por los tumores y además expresaban niveles bajísimos de PD-1, con lo que resultaban muy interesantes para fundar una nueva generación de inmunterapias contra tumores (Cruz-Adalia *et al.,* 2017).

Por lo tanto, es posible, haciendo uso de estos linfocitos entrenados por bacterias que expresan antígenos tumorales, lograr una respuesta inmunitaria mejorada frente a los tumores. Utilizando bacterias portadoras de neoantígenos que han sido capturadas y destruidas, por el proceso anteriormente descrito, por los linfocitos T CD4+ y actuando estos como APC para los linfocitos T CD8+, es posible originar linfocitos T CD8+ que darán lugar a una respuesta citotóxica (Cruz-Adalia et al., 2017). Además Todo este proceso induce la generación de memoria central con baja expresión de PD-1(Cruz-Adalia et al., 2017) lo cual, convierte a estos linfocitos en candidatos idóneos para utilizar inmunoterapia con neoantígenos.

Otra vía de aplicación de terapia inmunogénica es mediante la creación de vacunas personalizadas. Una vez se ha identificado el neoantígeno, se pueden crear vacunas personalizadas, que utilizadas junto con un adjuvante, pueden servir para generar una respuesta inmune en el individuo (Schumacher & Schreiber, 2015). Se sabe que la vacunación contra neoantígenos específicos de tumor pueden generar memoria minimizando la inducción potencial de tolerancia y el riesgo de autoinmunidad (Guo *et al.,* 2018).

Este tipo de vacunas ha demostrado un prometedor potencial terapéutico en ensayos preclínicos (Guo *et al.,* 2018).

Por ello, la clave para desarrollar, no solo esta nueva terapia antitumoral, sino todas las que están basadas en linfocitos, como las CAR-T cells, es desarrollar técnicas que permitan detectar moléculas expresadas de forma específica y excluyente en las células tumorales. Esta identificación de antígenos tumorales es el principal cuello de botella para avanzar en las terapias personalizadas. Pese a que los neoantígenos pueden identificarse mediante secuenciación, bioinformática y espectrometría de masas, identificar y distinguir aquellos que son inmunogénicos y capaces de promover el rechazo tumoral sigue siendo un desafío importante a día de hoy (Riley, 2019). Se pretende, como primer objetivo, desarrollar un método que permita la rápida y fácil identificación de posibles neoántigenos a partir de datos de secuenciacíon masiva (muy fáciles de obtener). Esto permitiría afrontar este cuello de botella de las inmunoterapias de una forma rápida y económica. Frente a los algoritmos informáticos clásicos utilizados para detectar neoantígenos, el uso de redes neuronales de machine learning (aprendizaje automatizado) puede resultar más potente y eficaz para poder conseguir un logro sustancial en este ámbito.

Hasta ahora, hay muy pocas investigaciones realizadas con redes neuronales para la deteccion de neoantígenos. No obstante es, desde hace tiempo usual utilizar redes neuronales para otras labores bioinformáticas como por ejemplo la predicción de la estructura secundaria de una proteina basándose en un su secuencia aminoacídica. De hecho ya se están utilizando algunos modelos de deep learning (aprendizaje profundo) para atacar estos problemas, concretamente se ha utilizado el modelo LSTM (Long Short Term Memory) para predecir la estructura secundaria con bastante éxito, 67,4 % de acierto (Sønderby, 2014). Este mismo modelo neuronal ya ha sido utilizado para la predicción de neoantígenos (Chen, 2019). Para la realización de este trabajo, se han utilizado modelos de LSTM y de redes neuronales convolucionales (CNN), para desarrollar programas que sean capaces de discernir patrones para la detección de neoantígenos capaces de causar respuesta antitumoral inmunitaria.

Las redes neuronales y concretamente, el aprendizaje automático, son modelos in-

formáticos que mediante la aportación de una serie de datos de entrenamiento, son capaces de establecer patrones y lograr "aprender" a resolver problemas en base a los los datos aportados. Las neuronas que componen las redes neuronales se basan en la regresión logísticas, es decir, el modelo básico de clasificación binaria. A continuación desarrollaremos brevemente el modelo de regresión logística por ser la base de todos los modelos de deep learning.

Un modelo de regresión logística intenta clasificar un punto de datos definidos con el vector x, cuyos valores pertenecen en general al conjunto de los números reales,

$$\mathbf{x} \in \mathbb{R}^n$$

en una de dos clases. Estas clases, para el problema que nos atañe serían no antigénico y sí antigénico, mientras que x sería una representación de la secuencia que estamos considerando. Para realizar la clasificación, el vector x n-dimensional es multiplicado por los pesos (conjunto de parámetros) que han sido calculados en el proceso de entrenamiento, o lo que es lo mismo, en el proceso de minimización de una función de error.

$$z = \mathbf{x} \cdot \underline{\omega} + b = \left(x^1, x^2, .., x^n\right) \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \\ \cdots \\ \omega_2 \end{pmatrix} + b = x^i \cdot \omega_i + b_{(i)}$$

donde

$$\underline{\omega} \in \mathbb{R}^n \qquad \text{y} \qquad b \in \mathbb{R}.$$

Al resultado de z, se le aplicará una función que transmite la información generada, de forma apropiada a las conexiones de salida, esta es la función de activación, que podrá ser una función rectificadora $ReLu(\zeta)$, una función escalón $u(\zeta)$ o la sigmoide $\sigma(\zeta)$ entre otras, o también se puede transmitir la información sin modificaciones. Cuando se llegue a la última capa, la función de activación dara el resultado esperado y'

$$ReLu(z) = \max\{0, z\}, \quad z \geq 0; \qquad u(z) = \begin{cases} 0 & \text{si } z < 0 \\ 1 & \text{si } z \geq 0 \end{cases}; \qquad \sigma(z) = \frac{1}{1 + e^{-z}}$$

Una vez entendido el funcionamiento de la regresión logística, entender el funcionamiento de una red neuronal es sencillo. Una red neuronal no es más que un grafo compuesto por capas de n neuronas. En las que cada neurona puede ser considerada un modelo de regresión logística. Una capa de entrada que recibe los datos de entrada, procesa los datos y transmite la información a la siguiente capa, llamada oculta, la cual procesa la información y la envía a la siguiente, hasta llegar a la última capa que es la que transmite la respuesta.

Independientemente de la complejidad del modelo, tanto una sencilla regresión logística como una complicada red neuronal se entrenan de igual forma. El proceso de entrenamiento consiste en el cálculo de unos pesos adecuados para que el sistema realice su labor de predicción de forma eficiente. El cálculos de los pesos está basado en el proceso de minimización de una función de error.

Esta red neuronal para que de resultados coherentes debe ser entrenada con un *dataset* de entrenamiento, es decir, un conjunto de datos que poseen los valores de entrada y los valores de salida que debe de dar la red. Los resultados deducidos por la red neuronal y' deben de ser lo más parecidos posibles al resultado del set de entrenamiento y. La función conocida como función de perdida $\mathcal{L}(y', y)$, mide la diferencia entre el valor esperado y y el valor obtenido y' por la red, para el caso sigmoide:

$$\mathcal{L}(y', y) = -(ylogy' + (1 - y)log(1 - y'))$$

Al poseer múltiples datos de entrenamiento, hay que vectorizar la función de pérdida. La vectorización de la función de perdida se llama función de coste y se define de la siguiente manera:

$$\mathcal{J}(\omega, b) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(y'_{(i)}, y_{(i)})$$

De nuevo, esta función nos da la diferencia entre los valores deducidos y los valores esperados. Como es lógico el mejor modelo, es decir, los pesos más idóneos, serán aquellos que hagan que la función de coste sea mínima. Por lo tanto para determinar

los pesos:

$$argmin\mathcal{J}(\omega, b) \Rightarrow (\omega, b).$$

En la fig. 4, se muestran dos redes neuronales que explica el funcionamiento de una red neuronal entrenada.

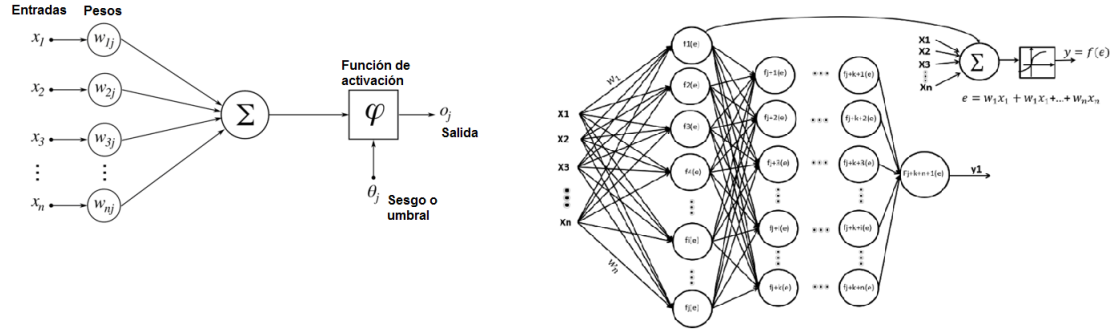

Figura 4: Dos redes neuronales, la primera sin una capa oculta y la segunda con varias capas ocultas. Fuente: https://www.researchgate.net/figure/Figura-1-Arquitectura-basica-de-una-red-neuronal-Haykin-1998-Partiendo-de-la-neurona_fig1_314151933

Con todo ello, las redes neuronales utilizadas en este trabajo han sido las redes neuronales convolucionales (CNN) y las redes neuronales LSTM (Long Short Term Memory). Las redes neuronales convolucionales están basadas en la convolución, que puede entenderse como una pequeña red neuronal clásica que se aplica parche a parche hasta cubrir todos los datos de entrada. Son un tipo de red neuronal que imita en cierta medida la organización y funcionamiento de las neuronas visuales. Consiste en utilizar varias capas ocultas de neuronas, unas conectadas a otras, de forma que cada capa extraiga características combinación de las caracterísiticas de capas previas, generándose una jerarquía de caracterísitcas de menor a mayor abstración (Lawrence, et al., 1997), de forma que cuantas más capas posea la red, más características podrá procesar esta. La otra red, la LSTM es un tipo de red neuronal recurrente (RNN). Una RNN, a diferencia de las redes neuronales no recurrentes, presenta memoria. En el caso más simple, una RNN formada por una única neurona genera una función en bucle que contiene la información precedente, así da el resultado en cada paso acorde a los resultados anteriores fig. 5.

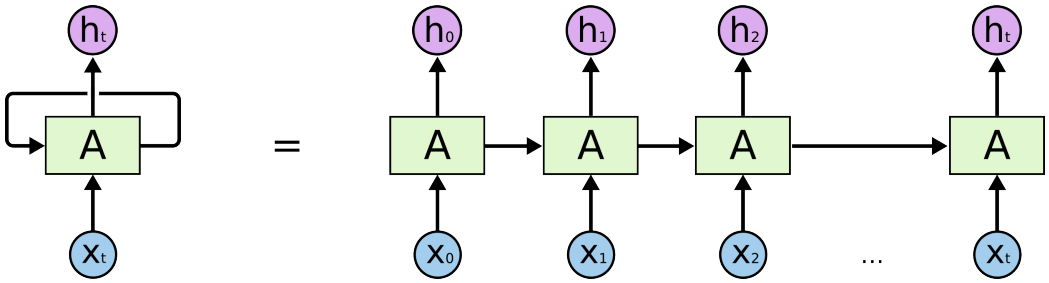

Figura 5: Funcionamiento de una RNN. Fuente: Olah, 2015

Lo mismo sucede cuando se crean capas con varias neuronas en una RNN, todas estas generan funciones internas que contienen la información anterior (Olah, 2015). Las redes LSTM son un tipo concreto de RNN, especialmente diseñadas para recordar información durante un periodo de tiempo superior, así si queremos traducir una secuencia, la LSTM no tendrá en cuenta sólo la última o últimas pocas letras, sino que tendrá en cuenta toda la información de la secuencia (Olah, 2015). Una RNN básica posee una única función de activación (tanh), mientras que la LSTM presenta una serie de cuatro funciones internas que operan sobre la información que reciben, tanto de la función interna como de la nueva entrada, haciendo así que una neurona pueda ir cambiando la información de salida según se avanza en la secuencia fig. 6.

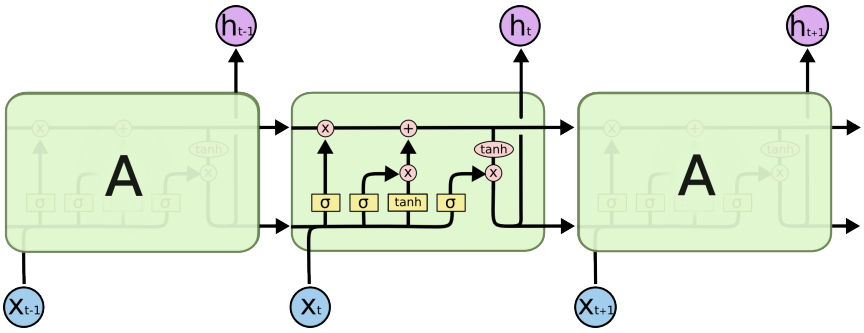

Figura 6: Funcionamiento de una LSTM. Fuente: Olah, 2015

Con todo esto, el objetivo de este trabajo es utilizando la base de datos de antígenos MHC I (IEDB.org), poder desarrollar modelos de redes neuronales convolucionales y LSTM que sean capaces de encontrar neoantigenos dentro de una muestra de RNA seq de un tejido tumoral. Esto supondría un método sistemático, rapido y sencillo para detectar neoantígenos de forma clínica, cosa que hoy en día no existe.

5. Resultados preliminares

5.1. Obtención del *dataset*

Se han obtenido varias bases de datos para entrenar los dos modelos de redes neuronales generados. para ello se partió de base de datos de IEDB (https://www.iedb.org/) de donde los datos extraidos fueron refinados y reducidos a aquellos péptidos antigénicos que poseían una respuesta alta o media, con un total de 112524 fragmentos peptídicos antigénicos. A partir de aquí se ha generado las bases de datos apropiadas para el entrenamiento de cada modelo. Para el modelo LSTM se generó una base de datos cuyas entradas son la secuencia de aminoácidos del fragmento antigénico y cuya salida es el fragmento antigénico que contiene o un 0 si no lo contiene. Para el caso de la convolucional, la entrada consiste en la secuencia aminoacídica y la salida en una secuencia de ceros y unos que corresponde a cada letra donde los 1 indicarían la presencia de un antígeno. Para ambas bases de datos se generaron distintas *dataset* de entrenamiento con distintas longitudes de las secuencias de entrada y de salida para observar con cual de ellas se entrenaban mejor las redes neuronales.

5.2. Resultados modelo Encoder-Decoder LSTM

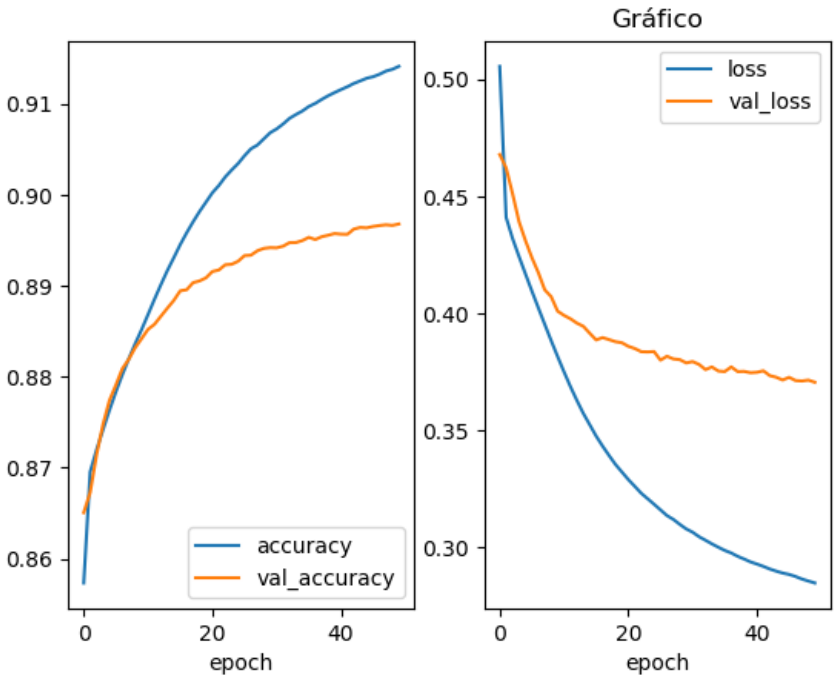

Figura 7: Resultados obtenidos del modelo RNN Encoder-Decoder. Donde en la subfigura de la izquierda se aprecia la evolución de la precisión (Accuracy) tanto para el caso del *dataset* de validación (val_accuracy) como para el *dataset* de entrenamiento (accuracy). En la subfigura de la derecha se aprecia la evolución de la pérdida (loss) de nuevo para el caso del *dataset* de validación (val_loss) y para el de entrenamiento (loss).

Como se aprecia en la figura 7 según avanzan las épocas de entrenamiento, mejora tanto el acierto y disminuye la pérdida. Entendiendose pérdida como el error que da la salida del sistema frente a la deseada.

Este resultado presenta un mejor valor de *loss* frente a otras que han sido calculadas para el desarrollo de este trabajo. Para lograr esta mejoría en el *loss* se han aumentado el numero de capas ocultas en el código de la red neuronal y también se ha aumentado del *dropout*, que viene a ser el silenciamiento de un porcentaje de neuronas en una capa dada para que el sistema no mecanice los resultados. También se modificó la base de datos con la que el sistema había sido redactado para explotar mejor la capacidad traductora de este tipo de redes. El mayor problema que presenta esta red es el *overfitting*, que se puede traducir como sobreajuste. El overfitting consiste en que la red "memoriza"patrones del conjunto de entrenamiento en vez de aprender a generalizar por tanto el aprendizaje automático falla. Para solventar este problema se probó a

reducir el numero de neuronas de las capas internas y a aumentar el dropout de las capas internas, para evitar el surgimiento de estos patrones. Realizado esto, en la red neuronal se redujo el valor de la pérdida y aumento el acierto, no obstante a partir de un punto no se lograba reducir la diferencia existente en la perdida del *dataset* de entrenamiento y el de validación.

Las medidas de acierto que aparecen en la figura 7 hacen referencia a la medida de acierto que hace la propia red neuronal, es decir, a los aciertos comparando letra por letra y también a la longitud de la secuencia traducida. Para tener una idea más intuitiva, más cercana a la clínica, la capacidad predictiva de la red a la hora de detectar antígenos en una secuencia, se estudió la capacidad que presentaba esta para dar lugar a un fragmento antigénico contenido en una secuencia de aminoacidos o la capacidad para dar negativo cuando este no estaba presente. Así se logró obtener que el acierto total de la red midiendo como acierto tanto a la predicción de negativos verdaderos y positivos verdaderos (*accuracy*) era de $65,6\,\%$. La sensibilidad, entendiéndose esta como la tasa de verdaderos positivos, es de $59,6\,\%$, la especificidad, que es la tasa de verdaderos negativos, $76,4\,\%$, la tasa de falsos positivos, $23,6\,\%$. Algunos de los resultados obtenidos fueron los mostrados en la tabla 1.

Tabla 1: Resultados obtenidos por la red LSTM en la predicción de antígenos

Secuencia peptídica	Antígeno	Predicción
MAAAAAGTATSQRFFQSFSDALIDEDPQ AALEELTKALEQKPDDAQYYCQRAYCHI	AAAAAGTATSQRF	AAAAAGTATSQR
MAAAAAAVGPGAGGAGSAVPGGAGPCAT SYIGEGAYGMVCSAYDNVNKVRVAIKKI	AAAAAAVGPGAGGAG	AAAAAAVGPGAGG
ATAATITTTMVAAAPVAVAAAAAPAAAA APSPATAAATAAAVSPAAAGQIPAAASV	AAAAPAAAA	AAAPAAAAA
AGRKTLRSCMGLEWFPELYPGYLGLGVL PGKPQCWNAMTQKPQLISPQGERLSQVS	AGRKT	0

5.3. Resultados modelo convolucional

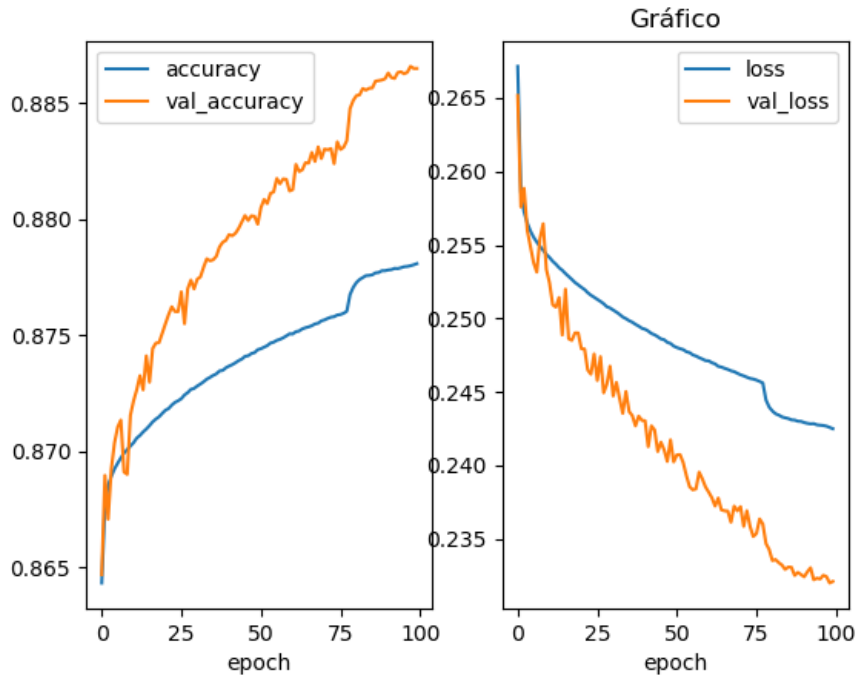

Figura 8: Resultados obtenidos del modelo convolucional unidimensional. Donde en la subfigura de la izquierda se aprecia la evolución de la precisión (Accuracy). En la subfigura de la derecha se aprecia la evolución de la pérdida (loss).

Como se aprecia en la figura 8 según se avanzan en el entrenamiento, se va mejorando la predicción y la pérdida va disminuyendo. Para lograr esta red neuronal se han añadido varias capas ocultas, para que mejorase el valor de la función de pérdida. Comparándola con otras que presentaban menos capas, esta última posee una perdida significativamente inferior y la precisión se ha mantenido pese a la adición de más capas ocultas.

A diferencia del caso anterior, apenas posee *overfitting* hecho que se ha logrado, fijando un valor de *dropout* bastante elevado en las capas internas, disminuyendo el numero de neuronas de las capas internas, y adaptando la base de datos para balancear los datos con los que se entrenaba la red.

De nuevo, la medición de acierto que aparecen en la figura 8 hacen referencia a la medida de acierto que hace la propia red neuronal. En este caso en cada cero y uno. Esta no es una forma apropiada de medir la precisión de este tipo de problemas, ya que

un antígeno de una única letra no tiene sentido. Por lo tanto estableciendo el criterio de que si considera tres positivos contiguos sí se considera antigénico se obtienen los siguientes resultados para la red neuronal: considerando de nuevo como acierto tanto la predicción de negativos verdaderos como de positivos verdaderos, se tiene que es de 83 %, la sensibilidad, es decir la tasa de aciertos positivos de 68, 5 %, la especificidad, tasa de verdaderos negativos, 95 % y la tasa de falsos positivos es baja 4, 1 %.

A continuación se muestran algunos de los resultados obtenidos por esta red.

Tabla 2: Resultados obtenidos por la red convolucional en la predicción de antígenos

Secuencia	GTRECSVTSDLDFPTQVIPLKTLNAVASVPIMYSWSPLQQNFMVEDETVLHNIPYM
Real	00000000000000000000000 01111111110000000000000000000000
Predicción	00000000000000000000000011111111100000000000000000000000
Secuencia	NLHACDSGASIRWQVVDRRIPLYANHKTLLQKAAAEFGAHY
Real	11110000000000000000000001111111111111111
Predicción	00000000000000000000000001111000111111111
Secuencia	+EAELQDMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYI
Real	11110000001111111111111110000011111111111111111111111111
Predicción	00000000001000001111111110000011111111111111111111111110
Secuencia	+VGKEDSSSTEFVEKRRAALERYLQRTVKHPTLLQDPDLRQFLESSELPRAVNTQAL
Real	00000111111111110000000000001111111111000000001111111111
Predicción	00000000000000000000000000000000000000000000001111000000
Secuencia	+HPPAPGMHPQAPGVHPQPPGVHPSAPGVHPQPPGVHPSNPGVHPPTPMPPMLRPPL
Real	00000000000000000000000000000000000000000011111111100000
Predicción	00000000000000000000000000000000000000000000000000000000

Como se aprecia en la tabla 2, La red neuronal tiende a dar a menudo falsos negativos, para poder solventar esto, una posible opción sería ver cuales son los pesos asociados a cada valor numérico en cada posición, y si fuera oportuno hacer más sensible la asignación del valor positivo. Otra posible opción para mejorar los resultados sería, una vez obtenidos los resultados, aplicar un algoritmo que donde hubiera pocos

ceros intercalando una secuencia de unos, convertirlos en unos, es decir, rellenar los huecos. Esto debería de realizarse para distintos valores de cantidad de ceros y unos y observar cual se ajusta mejor.
En ambos casos, para medir la eficiencia de la red neuronal se generó una base de datos nueva a partir de la base de datos original, así los fragmentos con los que se validaba no eran diferentes a los que había sido entrenada.

6. Plan de trabajo

La parte realizada *in silico* tenía la finalidad de lograr redes neuronales capaces de predecir mediante una secuencia de RNA-seq fragmentos neoantigénicos. Mientras que los experimentos a realizar en el laboratorio tienen como finalidad lograr dilucidar la efectividad que presentan las redes neuronales a la hora de predecir neoantígenos tumorales. Las secuencia de experimentos comienza con el uso de la red neuronal para predecir neoantígenos de un tipo de tumor específico dado y la síntesis de estos. Para después inyectar estos neoantígenos en ratones en forma de vacuna con un adyuvante para finalmente, llevar a cabo la implantación de tumores y realizar su seguimiento.

6.1. Generación de la *dataset* y desarrollo de las redes neuronales

Lo primero que hay que hacer para poder entrenar las redes neuronales es obtener una base de datos con la que la red pueda ser entrenada. Los datos de antígenos humanos expresados en el contexto de MHC I se obtuvieron a partir de la base de datos de IEDB (https://www.iedb.org/). Obtenidos estos junto con la referencia de la proteina a la que pertenecían, se desarrollo un algoritmo para obtener la secuencia FASTA de la proteína de la que provenía cada fragmento peptídico. Esta base de datos se adecuó de forma apropiada para entrenar las distintas redes neuronales. Para ello se tienen que establecer tanto ejemplos positivos como negativos. Para el caso de la convolucional, se desarrollo un algoritmo que fragmentaba las proteinas de forma apropiada y construía una base de datos donde los datos positivos consistían en una un fragmento de proteína, que sería la entrada, y su asociación a un código binario

compuesto por ceros y unos, que sería la salida. Los unos indicarían donde se encontraba la secuencia antigénica. Por el contrario, los fragmentos negativos se seleccionaron a partir de zonas de proteínas no catalogadas como antigénicas, consistitían en el fragmento de la proteína, como entrada, y toda una secuencia de ceros, como salida, indicando así la no presencia de antígeno. Para la LSTM, se desarrollo otro algoritmo que al igual que el anterior fragmentaba la proteína y de aquellos fragmentos que contenían alguna secuencia antigénica, dato de entrada, obtenían como dato de salida el propio antígeno, mientras que si la secuencia seleccionada no contenía ningún antígeno, la secuencia de salida sería un cero.
Para el desarrollo de las redes neuronales se utilizó la librería TENSORFLOW-KERAS de python. Que es una librería diseñada para el diseño de redes neuronales.

6.2. Predicción y síntesis del neoantígeno

Para la segunda parte, se pretendía validar un sistema para la predicción de neoantígenos en ratón que había sido diseñado en el laboratorio de acogida. En este caso se iba a utilizar un melanoma de ratón B16 como modelo tumoral para validar los algoritmos de predicción de neoantígenos. El primer paso es la síntesis de algunos de los neoantígenos asociados al tumor B16 murino con potencial para ser presentados por MHC-I, predichos por la red neuronal (tabla 3) (Wert *et al.,* 2019).

Tabla 3: Distintos candidatos a neoantígenos detectados por la red neuronal.

Nombre dado al péptido mutado	Secuencia aminoacídica(antígeno y mutación)
Pnp	SLITNKVVMEYENLEKANHM
Adar	LVPL SQAWTH PPGVVN PDSC
Nr1h2	VCGD KASGF R YNVLSC EGCK
Lrrc28	EPMF TFVYP T IFPLRE TPMA
Wiz	TASP PPTARM MFSGLA TPSL
Car11	LQGN FVPGP S FWGLVN AAWS
TPR2	PQIANCSVYDFFVWLHYYSV

Los peptidos predichos, Pnp, Adar, Lrrc28 y Wiz, se sintetizan en la unidad de proteómica del CNB-CSIC. Además el péptido TRP2 se sintetiza y se utiliza como

control positivo del experimento, ya que se trata de un neoantígeno conocido del tumor B16 murino que es presentado por MHC-I. Como control negativo se utilizaría NDVNAAIATIKTKRTI que es péptido de α-tubulina(323-337), péptido inerte que no genera respuesta inmune.

6.3. Preparación de los péptidos

El siguiente paso sería **preparar los péptidos**.

Los péptidos sintetizados por la unidad de proteómica son entregados liofilizados y se almacenan a 4ºC hasta que son resuspendidos en PBS a una concentración de 1 mg/ml. Hasta su uso, se conservan en alícuotas a $-80^{o}C$.Para realizar el experimento, a la dosis adecuada de cada péptido se le debe mezclar el adyuvante InjectAlum^{TM} (ThermoFisher Scientific) en una proporción de 1:1 (v/v) quedando una concentración resultante de $0,5mg/ml$. Esta mezcla se tiene que dejar media hora en agitación constante para lograr así una buena unión péptido-adyuvante que potenciará la respuesta inmune asociada al péptido.

6.4. Mantenimiento línea celular B16

Las células tumorales B16 F10 se criopreservan a $-198^{o}C$, en tanques de nitrógeno líquido. Para su uso se descongelan a $37^{o}C$ de forma rápida en un baño de agua y se cultivan en placas de cultivo celular de 1000 mm de diámetro en medio completo DMEM suplementado con 10-20 % FBS (Suero Bovino Fetal) y penicilina/estreptomicina.

Esta línea celular debe de mantenerse con pases de placa durante los 10 días precedentes a la implantación de tumores. Los pases deben realizarse cada dos días, las células B16, que son adherentes, se levantan con TRIPSINA-EDTA y se vuelven a cultivar en medio completo DMEM nuevo con las características anteriores.

6.5. Vacunación antitumoral con péptidos: diseño experimental

Para llevar a cabo el experimento se utilizarán un total de 42 ratones pertenecientes a la línea C57BL/6J wild type, con una edad promedio de 6-8 semanas todos del mismo sexo. Se repartirán en 8 grupos experimentales de la forma equitativa (6 ratones por grupo) tabla 4. Los grupos realizados fueron:

Tabla 4: Grupos experimentales.

Nombre del grupo	Características
α-tubulina	Control negativo (C-)
α-tubulina+adyuvante	Control negativo (C-)
TRP2+adyuvante	control positivo (C+)
Adar+adyuvante	Experimento
Wiz+adyuvante	Experimento
Pnp+adyuvante	Experimento
Lrrc28+adyuvante	Experimento
Adyuvante+Combinación de todos: TPR2, Adar, Wiz, Pnp, Lrrc28	Experimento

A cada ratón se le inyectan 50μg del péptido correspondiente en un volumen de 100μl, por dosis. Los ratones son vacunados con 3 dosis iguales, los días 0, 3 y 6 del experimento. A día 21 del experimento se implantan los tumores B16 de manera subcutánea: las células B16 se levantan y se cuentan con cámara Neubauer, y se inyectan en el flanco a una dosis de 400.000 células tumorales B16 por ratón con Matrigel Matrix (Corning) en una relación 1:1 (v/v) y un volumen final de 100 ul.

6.6. Implantación se seguimiento de tumores

Concluida la última dosis de vacunación con los neoantígenos, a día 21 del experimento se implantan los tumores B16 de manera subcutánea: como se ha comentado en el apartado anterior.

A partir de la inyección del tumor, día 21 del experimento, se deberá medir diariamente el tamaño tumoral de los 33 ratones con un calibre (Volumen=longitud·anchura$^2 \cdot \frac{1}{2}$)

para evaluar la eficacia de la vacunación en todos los grupos experimentales mediante el seguimiento de tumores, ya que nuestra medida de efectividad de la terapia es el crecimiento tumoral diferencial entre los distintos grupos experimentales.

En la figura 9 se muestra como sería el desarrollo experimental para el plazo de un año.

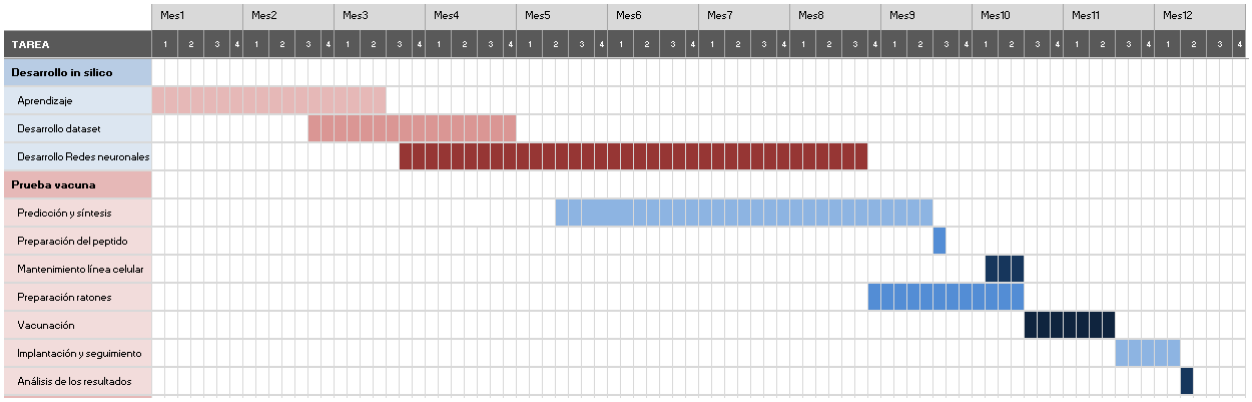

Figura 9: Cronograma que muestra el desarrollo experimental a lo largo del tiempo.

7. Bibligrafía

Anz, D., Eiber, S., Scholz, C., Endres, S., Kirchner, T., Bourquin, C., & Mayr, D. (2011). In breast cancer, a high ratio of tumour-infiltrating intraepithelial CD8+ to FoxP3+ cells is characteristic for the medullary subtype. Histopathology, 59(5), 965–974. https://doi.org/10.1111

Caldwell, C., Johnson, C. E., Balaji, V. N., Balaji, G. A., Hammer, R. D., & Kannan, R. (2017). Identification and validation of a PD-L1 binding peptide for determination of PDL1 expression in tumors. Scientific reports, 7(1), 1-11.

Chen, B., Khodadoust, M. S., Olsson, N., Wagar, L. E., Fast, E., Liu, C. L., ... & Davis, M. M. (2019). Predicting HLA class II antigen presentation through integrated deep learning. Nature biotechnology, 37(11), 1332-1343.

Cruz-Adalia, A., Ramirez-Santiago , G., Calabia-Linares , C., Torres-Torresano , M., Feo, L., Galán-Díez , M., . . . Veiga , E. (2014). T cells kill bacteria captured by transinfection from dendritic cells and confer protection in mice. Cell Host Microbe, 15(5), 611-622.

Cruz-Adalia, A., Ramirez-Santiago, G., Osuna-Pérez, J., Torres-Torresano, M., Zorita, V., Martínez-Riaño, A., ... & Alarcón, B. (2017). Conventional CD4+ T cells present bacterial antigens to induce cytotoxic and memory CD8+ T cell responses. Nature communications, 8(1), 1-11.

Cruz-Adalia, A., Ramírez-Santiago, G., Torres-Torresano, M., Garcia-Ferreras, R., & Chacón, E. V. (2016). T Cells Capture Bacteria by Transinfection from Dendritic Cells. JoVE (Journal of Visualized Experiments), (107), e52976.
1

Delves, P. J., Martin, S. J., Burton, D. R., Roitt, I. M., & Alonso, P. A. (2008). Roitt. Inmunología. Fundamentos (11a edición). Inmunología, 27(4), 212-214.

DeNardo, D. G., & Ruffell, B. (2019). Macrophages as regulators of tumour immunity and immunotherapy. Nature Reviews Immunology, 19(6), 369-382.

Desrichard, A., Snyder, A., & Chan, T. A. (2016). Cancer neoantigens and applications for immunotherapy. Clinical Cancer Research, 22(4), 807-812.

Farhood, B., Najafi, M., & Mortezaee, K. (2019). CD8+ cytotoxic T lymphocytes in cancer immunotherapy: A review. Journal of cellular physiology, 234(6), 8509-8521.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., ... & Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer, 136(5), E359-E386.

Guo, Y., Lei, K., & Tang, L. (2018). Neoantigen vaccine delivery for personalized anticancer immunotherapy. Frontiers in immunology, 9, 1499.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. cell, 144(5), 646-674.

Helmink, B. A., Reddy, S. M., Gao, J., Zhang, S., Basar, R., Thakur, R., ... & Gopalakrishnan, V. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. Nature, 577(7791), 549-555.

Ino, Y., Yamazaki-Itoh, R., Shimada, K., Iwasaki, M., Kosuge, T., Kanai, Y., & Hiraoka, N. (2013). Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. British journal of cancer, 108(4), 914–923. https://doi.org/10.1038/bjc.2013.

Kumar, H., Kawai, T., & Akira, S. (2011). Pathogen recognition by the innate immune system. International reviews of immunology, 30(1), 16-34.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks, 8(1), 98-113.

Netea, M. G., Joosten, L. A., Latz, E., Mills, K. H., Natoli, G., Stunnenberg, H. G., ... & Xavier, R. J. (2016). Trained immunity: a program of innate immune memory in health and disease. Science, 352(6284), aaf1098.

NIH: National Cancer Institute. (2020). Definición de inhibidor de puntos de control inmunitario. Diccionario de Cáncer. https://www.cancer.gov/espanol/publicaciones/diccionario/def/inhibidor-de-puntos-de-control-inmunitario

Olah, C. (2015). Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Riley, T. P., Keller, G. L., Smith, A., Devlin, J. R., Davancaze, L. M., Arbuiso, A. A., & Baker, B. M. (2019). Structure based prediction of neoantigen immunogenicity. Frontiers in immunology, 10, 2047.

SEOM: Sociedad Española de Oncología Médica. (2019). ¿Qué es la Medicina personalizada? . https://seom.org/informacion-sobre-el-cancer/ique-es-la-medicina-de-precision

Schreiber, R. D., Old, L. J., & Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science, 331(6024), 1565-1570.

Schumacher, T. N., & Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. Science, 348(6230), 69-74.

Shao, X. M., Bhattacharya, R., Huang, J., Sivakumar, I. A., Tokheim, C., Zheng, L., ... & Riemer, A. B. (2020). High-throughput prediction of MHC class i and ii neoantigens with MHCnuggets. Cancer Immunology Research, 8(3), 396-408.

Sønderby, S. K., & Winther, O. (2014). Protein secondary structure prediction with long short term memory networks. arXiv preprint arXiv:1412.7828.

Waldhauer, I., & Steinle, A. (2008). NK cells and cancer immunosurveillance. Oncogene, 27(45), 5932-5943.

Weinberg, R. A. (2014). The biology of cancer. 2nd edn. New York: Garland Science.

Wert, C., Muñoz, A. & Veiga, E. (2019). Murine pipeline for personalized anti-tumoural vaccines selection. Bachelor Thesis.

Wouters, M. C., & Nelson, B. H. (2018). Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer. Clinical Cancer Research, 24(24), 6125-6135.