



**Universidad
Europea MADRID**

***ESTUDIO DEL REPERTORIO DE
LINFOCITOS T DURANTE INMUNOTERAPIA
ADYUVANTE EN PACIENTES DE CÁNCER
DE PULMÓN DE CÉLULA NO PEQUEÑA
(NADIM)***

Autor: Cristina Martínez Toledo

Tutor del trabajo: Alberto Cruz Bermúdez

Tutor académico: María del Rocío González

***Facultad de Ciencias Biomédicas y de la Salud
Titulación: Máster Universitario en Bioinformática
Curso 2021-2022***

ÍNDICE:

ABSTRACT. PALABRAS CLAVE	2
GLOSARIO DE ABREVIATURAS	3
INTRODUCCIÓN	5
HIPÓTESIS Y OBJETIVOS	17
MATERIALES Y MÉTODOS	18
RESULTADOS	22
DISCUSIÓN	29
CONCLUSIONES	34
BIBLIOGRAFÍA	35
ANEXOS	39

ABSTRACT. PALABRAS CLAVE.

Introducción: La inmunoterapia como nuevo tratamiento para el cáncer de pulmón ha supuesto una gran mejora en la supervivencia de los pacientes metastáticos. Sin embargo, su respuesta no es igual en todos los pacientes y no se conoce bien su papel los estadios operables a través de tratamientos neoadyuvante (antes de la cirugía) o adyuvante, por lo que surge la necesidad de estudiar nuevos biomarcadores que permitan conocer mejor los mecanismos diferenciales entre los pacientes respondedores (CPR, respuesta patológica completa en la cirugía) y no respondedores (NCPR).

Materiales y Métodos: Se secuenciaron los repertorios de TCR de las muestras recogidas en sangre y tejido de los 46 pacientes participantes en el ensayo NADIM durante el tratamiento neoadyuvante y adyuvante. Una vez obtenidos los datos se procedió a un análisis de las métricas y el espacio clonal de sus clones más frecuentes. Además, se realizó un seguimiento de aquellos clones predominantes en las muestras de tejido.

Resultados: Los pacientes CPR, cuyos tumores responde de manera completa presentan unos repertorios menos uniformes ($p = 0,029$), diversos ($p = 0,025$) y con un mayor espacio clonal ocupado por su top 1% ($p = 0,025$) a los 6 meses de la adyuvancia en comparación con los NCPR. Pero los pacientes NCPR sufren cambios en las mismas métricas diferenciales, uniformidad ($p = 0,037$), diversidad ($p = 0,049$) y espacio clonal ($p = 0,020$) a los 12 meses que los equipara a los pacientes CPR. En cuanto a los seguimientos en sangre de los clones predominantes en tejido, los clones basales parecen ocupar un mayor espacio clonal en los pacientes CPR hasta llegar a ser significativamente distintos a los NCPR a los 12 meses de tratamiento ($p = 0,048$). Por otro lado, los clones post-neoadyuvancia pierden espacio clonal en los pacientes CPR durante la adyuvancia.

Conclusión: Aún tras la neoadyuvancia los pacientes CPR presentan una respuesta diferencial que parece indicar una selección y expansión clonal específica durante la adyuvancia, y que señala una promoción de los clones del tejido basal antes que los del tejido post-neoadyuvante. Este estudio describe posibles mecanismos de respuesta durante la adyuvancia que deberán ser estudiados más profundamente.

Palabras Clave: Cáncer de Pulmón, Inmunoterapia, TCR, Adyuvancia, Respuesta Patológica, Biomarcadores

GLOSARIO DE ABREVIATURAS:

- CD** – *Cluster* (Grupo) de Diferenciación
- CDR** – Complementary-Determining Region
- CPA** – Célula Presentadora de Antígeno
- CPCNP** – Cáncer de Pulmón de Células No Pequeñas
- CPCP** – Cáncer de Pulmón de Células Pequeñas
- CPR** – Complete Pathological Response
- CTLA4** – Cytotoxic T Lymphocyte Antigen 4
- EMA** – European Medicines Agency
- FDA** – Food and Drug Administration
- GCO** – Global Cancer Observatory
- HLA** – Human Leukocyte Antigen
- IASLC** – International Association for the Study of Lung Cancer
- ICI** – Immune Check-point Inhibitors
- ICOS** – Inducible T Cell CO-Stimulator
- IL-12** – Interleucina 12
- IFN- γ** – Interferón Gamma
- IPR** – Incomplete Pathological Response
- MHC** – Major Histocompatibility Complex
- MPR** – Major Pathological Response
- MRI** – Magnetic Resonance Imaging
- NCI** – National Cancer Institute
- NK** – Natural Killer
- OMS** – Organización Mundial de la Salud
- PET-CT** – Positron Emission Tomography-Computed Tomography
- PD1** – Programmed Death 1
- PD-L1** – Programmed Death-Ligand 1
- SEOM** – Sociedad Española de Oncología Médica
- SEER** – Surveillance, Epidemiology, and End Results Program
- TCR** – T Cell Receptor
- TMB** – Tumor Mutational Burden
- TNFR** – Tumor Necrosis Factor Receptor

TNM – Tumor primario / Nódulos linfáticos regionales afectados / Metástasis distante

INTRODUCCIÓN:

1. CÁNCER DE PULMÓN:

a) Cáncer. Situación Actual:

El cáncer, según la Organización Mundial de la Salud (OMS), se define como: “*Conjunto de enfermedades que se pueden originar en casi cualquier órgano o tejido del cuerpo cuando células anormales crecen de forma descontrolada, sobrepasan sus límites habituales e invaden partes adyacentes del cuerpo y/o se propagan a otros órganos.*” (OMS, 2022).

Actualmente, representa uno de los principales problemas de salud global y en España es una de las principales causas de morbi-mortalidad. El número de nuevos casos estimados para España en el año 2020 era de 282.000, pero se espera un incremento en los próximos años hasta llegar a los 375.000 en el año 2040 (SEOM, 2022; GCO, 2022).

b) Epidemiología:

El cáncer de pulmón es, de entre los diferentes tipos de cáncer, uno de los más incidentes a nivel mundial y, además, es aquel con más mortalidad (Fig. 1) (GCO, 2022).

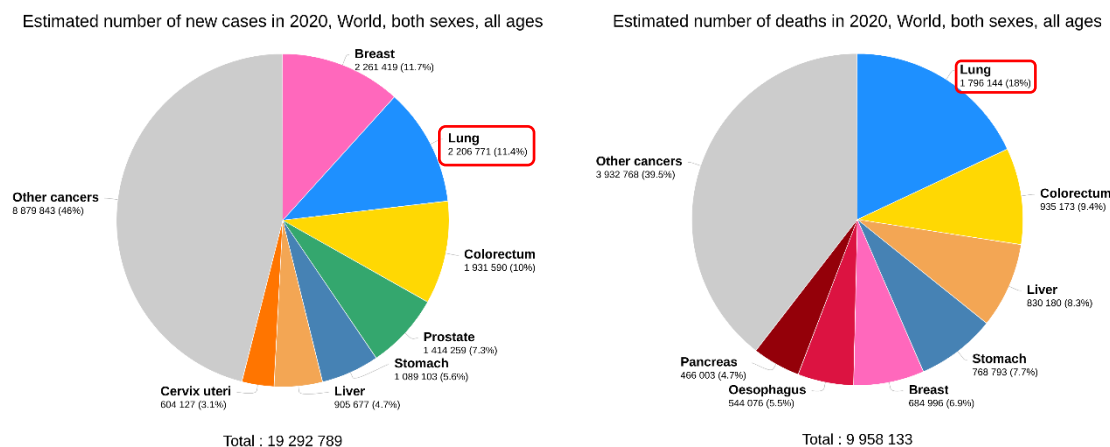


Figura 1. Número estimado de nuevos casos y muertes mundiales según el tipo de cáncer en el año 2020. (GCO, 2022)

En los últimos años, tras un largo estancamiento de la supervivencia de este cáncer, ha surgido una tendencia prometedora con un descenso significativo en el número de muertes (Fig. 2). Las principales causas que explican esta tendencia son: los avances en el diagnóstico y en procedimientos quirúrgicos; y la implementación de terapias dirigidas a mutaciones y la inmunoterapia como nuevas opciones de tratamiento. (Siegel et al., 2022)

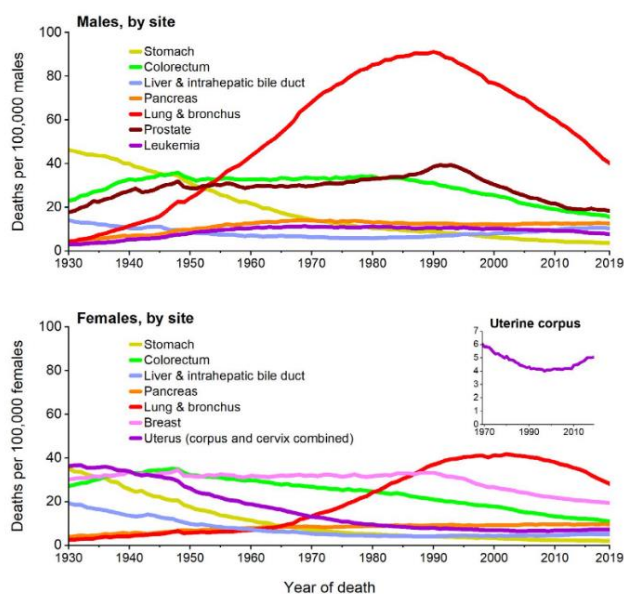


Figura 2. Tendencias en la mortalidad de determinados tipos de cáncer para ambos sexos en Estados Unidos (1930-2019). (Siegel et al., 2022)

c) Clasificación y Estadificación:

El cáncer de pulmón se clasifica histológicamente en dos grandes grupos: el cáncer de células no pequeñas (CPCNP), el mayoritario con un 85% de los casos; y el cáncer de célula pequeña (CPCP) con un 15% de los casos.

El CPCNP se subdivide a su vez, en otras tres categorías: el adenocarcinoma (40% de todos los casos de cáncer de pulmón), carcinoma escamoso (25-30%) y el de célula grande (5-10%).

Sin embargo, a pesar de su diversidad histológica, el tratamiento escogido es dependiente de su estadio que proporciona información sobre la extensión anatómica del cáncer. Actualmente el proceso de estadificación se realiza siguiendo las pautas de la clasificación TNM propuesta por la International Association for the Study of Lung Cancer (IASLC) (Detterbeck et al., 2017; Duma et al., 2019).

Los parámetros utilizados para realizar esta clasificación son los siguientes:

- **T (Tumor Primario):** Determinado por el tamaño del tumor, la invasión adyacente central/mediastínica o de las estructuras periféricas y la distancia de un nódulo tumoral adicional al tumor primario. En base a esta información es capaz de clasificar al tumor en 5 categorías (T0 a T4).
- **N (Nódulos Linfáticos Regionales Afectados):** Como su nombre indica informa de la localización de nódulos involucrados o afectados por el tumor para su clasificación en 4 categorías (N0 a N3).
- **M (Metástasis Distante):** Informa de la existencia de nódulos tumorales adicionales en las estructuras pleurales o pericárdicas (categoría M1a), la existencia de una sola metástasis distante o extratorácica (categoría M1b); o la aparición de múltiples metástasis, ya sea en uno o varios órganos (categoría M1c).

Una vez recopilada toda esta información y según las categorías seleccionadas para cada parámetro, al tumor podrá asignársele un estadio determinado (Fig. 4) (Detterbeck et al., 2017).

A		B					
T (Primary Tumor)	Label	T/M	Subcategory	N0	N1	N2	N3
T0	No primary tumor	T1	T1a	IA1	IIB	IIIA	IIIB
Tis	Carcinoma in situ (Squamous or Adenocarcinoma)		T1b	IA2	IIB	IIIA	IIIB
			T1c	IA3	IIB	IIIA	IIIB
T1	Tumor ≤3 cm,	T2	T2a	IB	IIB	IIIA	IIIB
T1a(mi)	Minimally Invasive Adenocarcinoma		T2b	IIA	IIB	IIIA	IIIB
T1a	Superficial spreading tumor in central airways ^a	T3	T3	IIB	IIIA	IIIB	IIIC
T1a	Tumor ≤1 cm		T4	IIIA	IIIA	IIIB	IIIC
T1b	Tumor >1 but ≤2 cm		M1	M1a	IVA	IVA	IVA
T1c	Tumor >2 but ≤3 cm		M1b	IVA	IVA	IVA	IVA
			M1c	IVB	IVB	IVB	IVB
T2	Tumor >3 but ≤5 cm or tumor involving: visceral pleura ^b , main bronchus (not carina), atelectasis to hilum ^b						
T2a	Tumor >3 but ≤4 cm						
T2b	Tumor >4 but ≤5 cm						
T3	Tumor >5 but ≤7 cm or invading chest wall, pericardium, phrenic nerve or separate tumor nodule(s) in the same lobe						
T4	Tumor >7 cm or tumor invading: mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus, spine; or tumor nodule(s) in a different ipsilateral lobe						
N (Regional Lymph Nodes)							
N0	No regional node metastasis						
N1	Metastasis in ipsilateral pulmonary or hilar nodes						
N2	Metastasis in ipsilateral mediastinal/subcarinal nodes						
N3	Metastasis in contralateral mediastinal/hilar, or supraclavicular nodes						
M (Distant Metastasis)							
M0	No distant metastasis						
M1a	Malignant pleural/pericardial effusion ^c or pleural/pericardial nodules or separate tumor nodule(s) in a contralateral lobe;						
M1b	Single extrathoracic metastasis						
M1c	Multiple extrathoracic metastases (1 or >1 organ)						

TX, NX: T or N status not able to be assessed

^a Superficial spreading tumor of any size but confined to the tracheal or bronchial wall

^b such tumors are classified as T2a if >3≤4 cm, T2b if >4≤5 cm.

^c Pleural effusions are excluded that are cytologically negative, non-bloody, transudative, and clinically judged not to be due to cancer.

Figura 4. Tablas guía para la estadificación en cáncer de pulmón según la 8ª edición de la clasificación TNM propuesta por la IASLC. (A) Descripción de las diferentes categorías para TNM (Detterbeck et al., 2017) (B) Guía para la estadificación según los parámetros TNM (Detterbeck, 2018).

d) Supervivencia:

La supervivencia del cáncer del pulmón, incluso con los últimos avances ya mencionados, sigue siendo una de las más bajas (22% a los 5 años) en comparación a otros tipos de cáncer como el de próstata (98%) o el melanoma (93%). Esta supervivencia está determinada por varios factores como el sexo o la edad, pero, sobre todo, el estadio en el que son diagnosticados los pacientes (Fig. 5) (Siegel et al., 2022).

Como se puede ver reflejado en la figura 5, la supervivencia es bastante mejor en aquellos estadios más localizados (~60%), pero, normalmente este cáncer es diagnosticado en una fase avanzada o distante, la cual tiene muy baja tasa de supervivencia (~7%). Esto explicaría la baja supervivencia general para este cáncer (NCI, 2022).

El diagnóstico actual no se realiza normalmente hasta una fase avanzada porque las técnicas implementadas hoy en día (diferentes tipo de toma de imágenes y evaluación patológica de biopsias) no son suficientes para detectar el desarrollo de cáncer en su fase temprana (Nooreldeen & Bach, 2021).

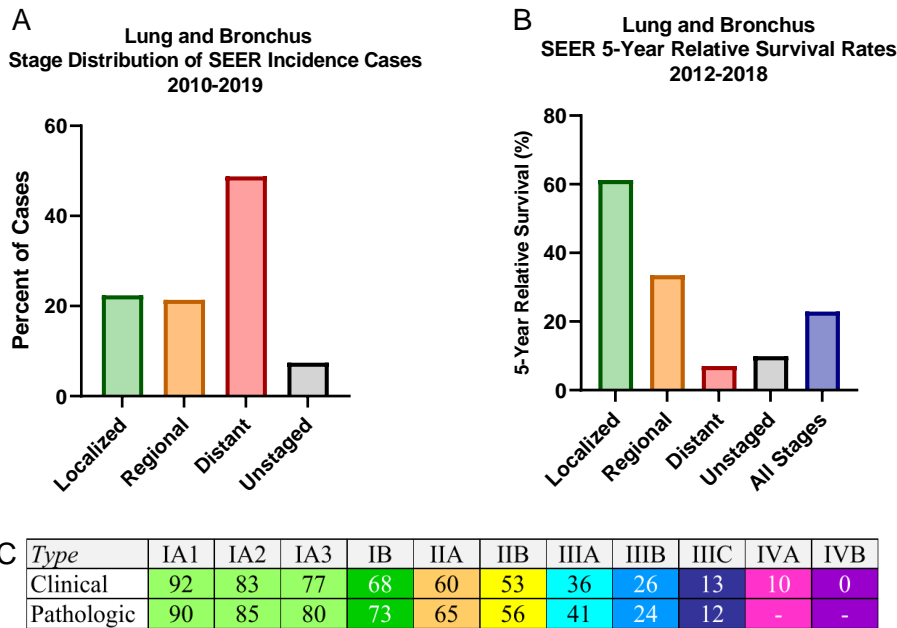


Figura 5. Gráficas sobre la incidencia y supervivencia en cáncer de pulmón según el estadio*. (A) Incidencia de casos de cáncer de pulmón según el estadio entre los años 2010 y 2019. (B) Supervivencia relativa a los 5 años según el estadio en el momento del diagnóstico entre los años 2012 y 2018. (C) Tabla con las supervivencias a los 5 años según el estadio (clasificado según TNM) para dos momentos, el clínico o antes del inicio del tratamiento; y el patológico, recogido tras una resección quirúrgica. (NCI, 2022; Detterbeck et al., 2017).

*Gráficas realizadas en GraphPad Prism 9 con datos del Surveillance, Epidemiology, and End Results (SEER) program por el National Cancer Institute (NCI) de Estados Unidos.

e) Tratamientos:

Como se mencionó previamente el tratamiento escogido es dependiente del estadio en el que se encuentre el paciente.

En aquellos con estadios más tempranos o localizados (I o II) se recomienda una resección completa del tumor por medio de diferentes técnicas (lobectomía, segmentectomía, etc.). De entre todas estas técnicas se escoge aquella que sea más adecuada para cada paciente según las limitaciones, complicaciones y beneficios de cada una. Además, se recomienda la administración de una quimioterapia adyuvante (o después de la cirugía), sobre todo, en aquellos casos con un gran tamaño tumoral. Esto es debido a que aumenta significativamente la supervivencia de estos pacientes y que las metástasis distantes son la principal causa de recaída tras una cirugía potencialmente curativa (Duma et al., 2019; Hoy et al., 2019).

Por otro lado, el estadio III o localmente avanzado es bastante heterogéneo ya que comprende casos reseccables, potencialmente reseccables o no operables, y cuyo tratamiento debe ser valorado por un equipo multidisciplinar. También, es fundamental un estadiaje preciso que de información sobre la extensión del tumor. Para ello, se procede a realizar una imagen PET-CT y MRI para detectar metástasis craneales o extracraneales (Calvo et al., 2021; Postmus et al., 2017).

- **Estadío III Potencialmente Resecable:** Su tratamiento es similar al de los estadios iniciales con una cirugía radical seguido de sesiones de quimioterapia adyuvante. Sin embargo, la supervivencia ha permanecido baja y por ello se proponen el uso de nuevos tratamientos como la inmunoterapia basada en inhibidores de puntos de control del sistema inmune (ICI, *Immune Check-point Inhibitors*). Aunque la quimioterapia neoadyuvante no ha sido tan evaluada como la adyuvante, su incorporación al tratamiento anterior supone una serie de ventajas como: identificar cuáles son los pacientes que responden, tratamiento temprano de las micrometástasis o bajada del estadio para una mejor resecabilidad del tumor.
- **Estadío III No Resecable:** Tradicionalmente estos casos eran tratados con una terapia combinada de quimioterapia y radioterapia, que mostraban mejores resultados que administradas de forma individual. Sin embargo, el tratamiento actual está basado en una quimioterapia acompañada de un posterior tratamiento inmune basado en anticuerpos contra el ligando de muerte programada 1 (PD-L1, *Program Death Ligand 1*). Gracias a las mejoras en la supervivencia que supuso este cambio, surgió el interés por la introducción de la inmunoterapia en combinación de otras modalidades de tratamiento (Calvo et al., 2021).

Por último, los pacientes con estadios avanzados (IV) se les trataba tradicionalmente con quimioterapia, pero la introducción de la inmunoterapia en este tipo de casos ha cambiado el panorama del tratamiento de cáncer de pulmón (Duma et al., 2019).

2. INMUNOTERAPIA:

a) Inmunoedición:

La inmunoterapia es un tipo de tratamiento que utiliza el sistema inmune del paciente, para que este sea capaz de reconocer y atacar a las células cancerosas. Esta capacidad de poder ejercer una respuesta efectiva contra el cáncer es algo que se pierde a través de la inmunoedición. Además, este proceso de adquisición de mecanismos de evasión del sistema inmune fue incorporado en 2011 como uno de los sellos característicos o “*Hallmarks*” del cáncer (American Cancer Society, 2022; Hanahan & Weinberg, 2011).

El proceso de inmunoedición divide la interacción del sistema inmune y las células tumorales en tres etapas diferentes (Fig. 6):

- **Eliminación o Inmunovigilancia:** Es aquella fase donde el sistema inmune (innato y adaptativo) es capaz de reconocer y eliminar aquellas células tumorales que han escapado los controles de supresión tumoral (ej. p53). Este ataque al tumor funciona por el proceso siguiente:
La masa tumoral al crecer de forma invasiva provoca disrupciones en los tejidos cercanos, y ello genera a su vez una respuesta inflamatoria. Esta respuesta recluta a las células del sistema inmune innato (ej. células Natural Killer o NK) y las estimula para que produzcan IFN- γ . Esta citoquina puede inducir la muerte tumoral de forma limitada por mecanismos antiproliferativos y apoptóticos. Además, puede estimular la producción de más citoquinas desde tanto las células cancerosas como las sanas de tejidos adyacentes.

Algunas de estas nuevas citoquinas pueden tener actividad angiostática por lo que fomentarían, aunque de forma indirecta, la muerte de las células cancerosas. El resto de ellas se encargarán de reclutar más células NK y macrófagos.

Las células NK y macrófagos que consiguen infiltrarse en el tumor eliminan las células tumorales por diferentes mecanismos, como la apoptosis dependiente de ligando o perforina. Además, son capaces de estimularse mutuamente mediante la producción de IFN- γ e IL-12.

Los restos de las células cancerosas ya muertas serán digeridos y transportados a los nódulos linfáticos por las células dendríticas. Una vez en los nódulos, se fomentará la generación de las células T CD4+ y T CD8+ específicas. Estas células serán transportadas al tumor donde las CD8+ destruirán las células cancerosas con antígeno (cuya inmunogenicidad es fomentada una vez más por el IFN- γ) (Dunn et al., 2002; Hong et al., 2014; O'Donnell et al., 2019).

- **Equilibrio:** Esta fase se produce en el momento que alguna de las variantes celulares que genera el tumor es capaz de escapar al ataque del sistema inmune. Es entonces cuando inicia un crecimiento de forma limitada sin provocar manifestaciones clínicas detectables en el paciente. Es la fase más larga, llegando a durar incluso años.

Esta producción de múltiples variantes es causa de la inestabilidad genética de las células cancerosas y para contrarrestar la presión a la que son sometidas debido al ataque del sistema inmune. Posteriormente, las variantes son seleccionadas hasta quedar aquellas que presentan una menor inmunogenicidad y por ello, cuentan con una menor posibilidad de ser eliminadas por la respuesta inmune. Algunos ejemplos de estas modificaciones que sufren las células cancerosas puede ser la pérdida de neoantígenos (péptidos que son reconocidos por las células inmunes) o la pérdida de la estabilidad del complejo de histocompatibilidad o MHC (*Major Histocompatibility Complex*) (Dunn et al., 2002; O'Donnell et al., 2019; Rosenthal et al., 2019).

- **Escape:** Ocurre cuando finalmente las variantes resistentes consiguen crecer de forma descontrolada y la enfermedad comienza a presentar manifestaciones clínicas (Dunn et al., 2002; O'Donnell et al., 2019).

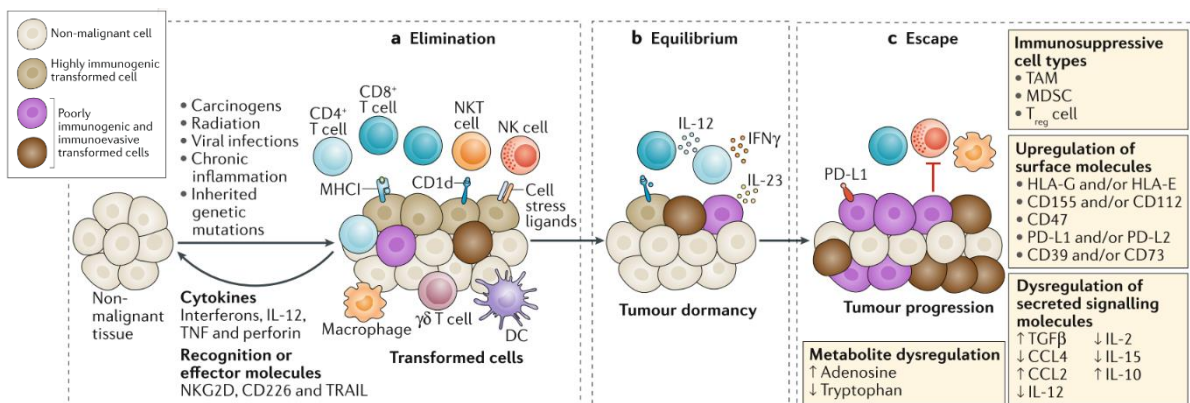


Figura 6. Esquema con los mecanismos moleculares durante las tres fases (eliminación, equilibrio y escape) de la inmunoección del cáncer. (O'Donnell et al., 2019)

b) Linfocitos T:

Los linfocitos son las células representantes del sistema inmune adaptativo, que se encarga de reconocer los péptidos ajenos o antígenos y generar una respuesta efectiva contra ellos que se pueda repetir en sucesivos contactos. Dentro de esta inmunidad adaptativa, hay dos tipos de respuesta: la humoral de la que se encargan los linfocitos B (producción de anticuerpos específicos tras reconocer el antígeno) y la celular en la que participan los linfocitos T (reacción celular tras el reconocimiento del antígeno).

Los linfocitos T poseen una doble especificidad en el reconocimiento de antígenos. Por un lado, son capaces de reconocer el MHC propio de las células y, además, también reconocen el antígeno que presenta este MHC de las células presentadoras de antígeno (CPA). Este reconocimiento del linfocito T se produce gracias al receptor de linfocitos T (TCR, *T Cell Receptor*) (Abbas et al., 2022).

Receptor de Células T (TCR):

Es un receptor con una distribución clonal, es decir, que cada clon de célula T posee un receptor distinto que es capaz de reconocer un antígeno determinado. Su estructura consiste en un heterodímero formado por dos cadenas (α y β) unidas por puentes disulfuro. Cada una de estas cadenas está formada por dos dominios, variable y constante (cuyos genes sufren una recombinación somática durante la maduración del linfocito T en el timo); una región hidrofóbica transmembrana y una pequeña región citoplasmática.

El dominio variable, a su vez, está compuesto por las regiones variable, diversidad y unión (esta última solo presente en la cadena β). Este dominio a su vez es el responsable del reconocimiento del complejo MHC-antígeno y es donde se acumula la mayor variabilidad entre los diferentes receptores y donde se encuentran las regiones determinantes de complementariedad (CDR, *Complementarity-Determining Regions*) (Fig. 7) (Abbas et al., 2022; Turner et al., 2006).

El receptor está asociado, además, a las proteínas ζ y CD3 de forma no covalente para formar el complejo TCR. Estas proteínas se encargan de transducir la señal que genera la activación de la célula T, esto es debido a que las regiones citoplasmáticas del heterodímero son demasiado cortas como para llevar a cabo esta tarea (Abbas et al., 2022).

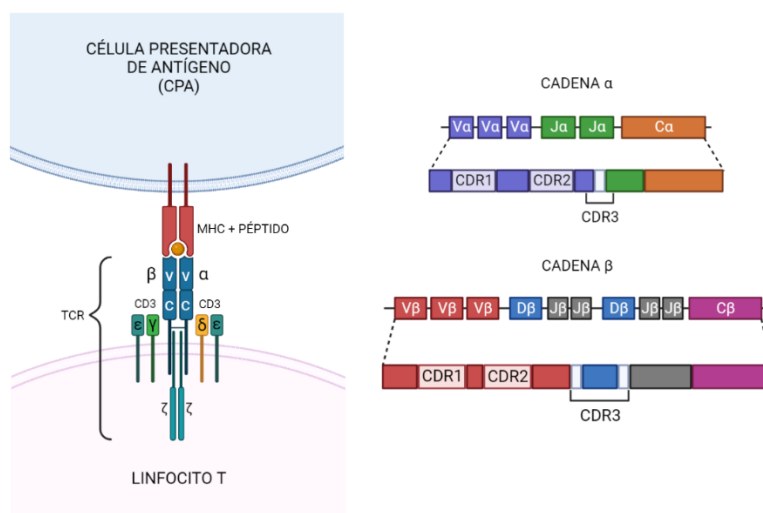


Figura 7. Estructura del complejo TCR y generación de las cadenas de TCR mediante recombinación somática de sus segmentos de genes.

**Figura de elaboración propia en base a la bibliografía consultada.*

Activación de los Linfocitos T:

Como se ha mencionado anteriormente, los linfocitos son activados al reconocer el complejo MHC-antígeno presentado por las CPA (ej. células dendríticas). Esta activación se da de dos formas según la madurez de linfocito (definida por la exposición previa al antígeno). Si el antígeno activa a un linfocito virgen, este dará lugar a una proliferación o expansión clonal que dará paso a células efectoras o de memoria. Estas células efectoras o maduras al ser de nuevo activadas por el antígeno serán activadas, pero en este caso, para ejercer su función que conllevará la eliminación de la fuente del antígeno.

Es necesario tener en cuenta que, también, según el subtipo de linfocito T que sea llevará a cabo una función u otra. Los linfocitos T colaboradores o CD4+ se encargan de la activación de otros tipos celulares como los linfocitos B, los linfocitos T CD8+ o los macrófagos. Por otro lado, los linfocitos T citotóxicos o CD8+ tienen una acción apoptótica contra las células infectadas o tumorales (Abbas et al., 2022).

Reconocimiento de las Células Tumorales:

Este reconocimiento se produce de la siguiente manera, los péptidos presentados por las células tumorales a través del complejo MHC son reconocidos por el TCR de los linfocitos T, lo que provoca su activación. Sin embargo, para que esta activación se produzca es necesaria una segunda señal estimuladora que se lleva a cabo por la unión del correceptor CD28 con los ligandos CD80 o CD86 que presenta la célula presentadora (célula tumoral en este caso). Posteriormente, el linfocito comienza a expresar receptores inhibitorios como el de muerte programada 1 (PD1, *Programmed Death 1*) como forma de terminar la respuesta una vez eliminada la célula tumoral (Fig. 8) (Havel et al., 2019).

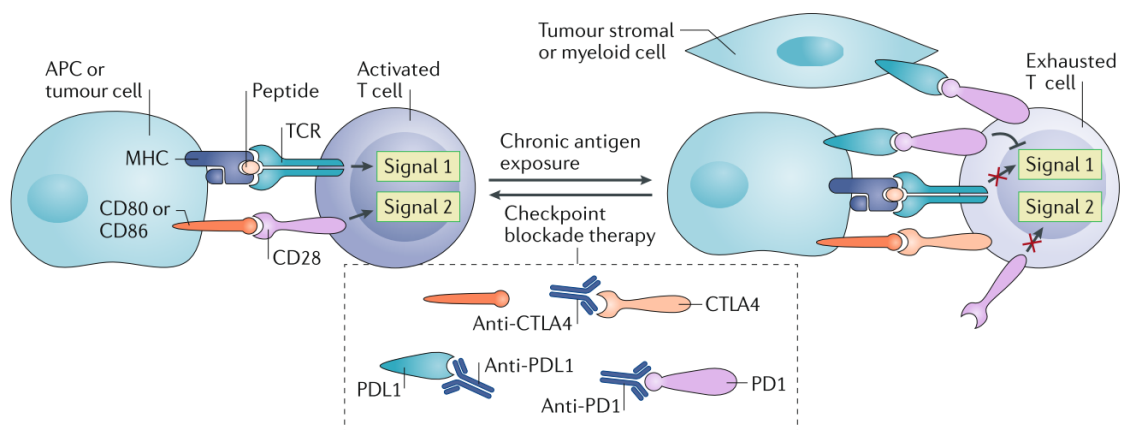


Figura 8. Reconocimiento del complejo antígeno-MHC y activación de los linfocitos T. (Havel et al., 2019)

Por lo tanto, este proceso es llevado a cabo por varios elementos además del TCR:

- **Antígenos Tumorales:**
 - **Neoantígenos:** Son aquellos antígenos producidos por las células cancerígenas que pueden ser detectados por los linfocitos T. Su característica principal es que son péptidos que no forman parte del genoma humano y, por lo tanto, son creados a partir de mutaciones en el ADN de las células tumorales.

- **Antígenos “propios”:** Péptidos que forman parte del genoma humano, pero que no poseen una tolerancia tan restrictiva debido a su baja expresión en determinados tejidos. De esta manera su sobreexpresión aberrante en el tumor también desencadenaría una respuesta por parte de los linfocitos T (Schumacher & Schreiber, 2015).
- **MHC:** La función de estas moléculas es la de unirse a péptidos para formar el complejo MHC-antígeno, que este pueda ser reconocido por los linfocitos T y con ello generar una respuesta inmune. Existen dos principales clases:
 - **MHC Clase I:** Es expresado por casi todas las células nucleadas (incluidas las células tumorales) y puede unirse a péptidos pequeños de proteínas intracelulares para ser presentados a los linfocitos T CD8+
 - **MHC Clase II:** Lo expresan principalmente las CPA y los linfocitos B. Se encargan de presentar péptidos (más grandes que los de clase I) de proteínas extracelulares a los linfocitos T CD4+.

Está codificado por seis genes del antígeno leucocitario humano (HLA, *Human Leukocyte Antigen*) y es bastante polimórfico lo que permite la unión a una gran variedad de péptidos. Su expresión en la zona tumoral puede condicionar la activación e infiltración de linfocitos T (Abbas et al., 2022; Havel et al., 2019; Perea et al., 2017; Trowsdale, 2005).

- **Moléculas Co-Estimuladoras:** Son aquellas que proporcionan la segunda señal necesaria para la activación del linfocito T. Destacan CD28 e ICOS (*Inducible T Cell CO-Stimulator*) con la unión de sus ligandos B7-H2, B7-1 y 2 (Chen & Flies, 2013).
- **Moléculas Inhibitorias:** Necesarios para limitar la actividad de los linfocitos una vez ejecutada su respuesta, por ello también son conocidos como puntos de control del sistema inmune. Algunos de ellos son el CTLA4 (*Cytotoxic T Lymphocyte Antigen 4*), o el PD1 que al unirse con ligando PD-L1 desencadena una cascada de señales que provocan una disminución de la señalización del TCR, de las moléculas co-estimuladoras y de la proliferación y supervivencia de los linfocitos T (Sharpe & Pauken, 2018).

c) Inmunoterapia y Limitaciones:

Como se ha introducido anteriormente, la inmunoterapia busca dirigir la respuesta inmunológica hacia la eliminación de las células tumorales. Por lo tanto, se utilizan agentes que promuevan esa respuesta antes evadida por el tumor (durante las fases de equilibrio y escape de la inmunoección).

Los inhibidores de puntos de control (ICIs) son la clase de inmunoterapia más estudiada a día de hoy, siendo los más importante los anticuerpos dirigidos al bloqueo de PD1 / PD-L1 y CTLA4. A pesar de que en condiciones fisiológicas estas moléculas funcionan como forma de controlar la respuesta linfocitaria, las células tumorales sobreexpresan sus ligandos para bloquear el ataque por parte de los linfocitos T y escapar de la respuesta inmune. Por ello, el uso de estos anticuerpos revierte este bloqueo para reactivar y fomentar el ataque a las células tumorales.

Sin embargo, también poseen sus limitaciones como sus efectos secundarios y la falta de respuesta en muchos pacientes. Y esto representa las principales limitaciones de la inmunoterapia. Su respuesta es muy diferente y variada entre los pacientes ya que no todos se benefician ni responden igual, y también ni presentan los mismos efectos adversos. Por ello, surge una necesidad de desarrollar biomarcadores para mejorar la identificación de pacientes que se benefician de esta terapia y evitar a pacientes no respondedores efectos secundarios no necesarios (Riley et al., 2019).

d) Biomarcadores:

Un biomarcador, según el Instituto Nacional del Cáncer (NCI, *National Cancer Institute*) es: “*Molécula biológica encontrada en sangre, otros fluidos corporales o en tejidos que funciona como señal de un proceso normal o anormal, de una enfermedad o de la respuesta a un tratamiento*”. Existen varias categorías de biomarcadores según el objetivo para el que se vayan a utilizar:

- **Biomarcador Diagnóstico:** Sirven como ayuda a la hora de emitir un diagnóstico diferencial (elegir un diagnóstico entre varios posibles) que viene acompañado de una serie de síntomas. Muchas veces es el parámetro determinante.
- **Biomarcador Pronóstico:** Informan sobre el desarrollo de la enfermedad en el paciente (supervivencia, capacidad de recuperación) de forma independiente a la administración de cualquier tratamiento. Se recogen en el momento del diagnóstico.
- **Biomarcador Predictivo:** También se recoge previo al tratamiento. En este caso informa sobre un resultado clínico (supervivencia, respuesta patológica) de forma dependiente a la administración de un fármaco o terapia determinada. Sirven para optimizar tratamientos e identificar a pacientes respondedores.
- **Biomarcador de Monitorización:** Utilizados para el seguimiento de la respuesta o eficacia de una intervención a lo largo del tiempo.

Ejemplos de Biomarcadores Utilizados en Inmunoterapia:

- **Expresión de PD-L1:** La determinación de la expresión de PD-L1 por inmunohistoquímica es un requisito necesario para la administración de Pembrolizumab (anticuerpo contra PD1) según la ficha técnica del fármaco elaborada por la Agencia Europea del Medicamento (EMA, *European Medicines Agency*). Es una suposición lógica que para que el medicamento haga su efecto, debe haber una expresión suficiente de su molécula diana (o ligando), y hay varios estudios que lo apoyan como el KEYNOTE-001. En este estudio, se observaron que aquellos pacientes con al menos un 50% de expresión de PD-L1 poseían mayores posibilidades de responder a la terapia (45,2%) que aquellos con una expresión menor (16,5%). Sin embargo, no es un biomarcador con una capacidad de identificación ideal. Existen un número importante de pacientes con una expresión alta de esta molécula que no responden al tratamiento, y también pacientes con expresiones consideradas como bajas que aún pueden responder.

La explicación de esta imprecisión puede estar en que la expresión de este ligando es muy variable tanto espacial como temporalmente. Por lo tanto, la toma de una muestra de una zona y momento determinado no representa de forma óptima la situación tumoral (Calvo et al., 2021; EMA, 2015; Garon et al., 2015; Gibney et al., 2016).

- **Linfocitos Infiltrantes al Tumor:** Su presencia está asociada a una mayor supervivencia en varios tipos de cáncer. Centrándose en CPCNP, una densidad alta de linfocitos T CD8+ en el tumor se relaciona a una mayor supervivencia general y libre de progresión. Podría explicarse debido a que el reconocimiento linfocitario de los tumores desembocaría una respuesta inflamatoria ligada a los linfocitos T y el pronóstico de la enfermedad sería mejor. Sin embargo, presenta los mismos problemas que el biomarcador anterior y falla aún de forma considerable (Gibney et al., 2016; Tokito et al., 2016).
- **Carga Mutacional Tumoral (TMB, *Tumoral Mutational Burden*):** Las mutaciones de las células tumorales son las responsables de la aparición de los neoantígenos que pueden ser reconocidos por el receptor de los linfocitos T y desencadenar una respuesta antitumoral. Por lo tanto, un mayor TMB se asocia a una mayor respuesta inmune y mayor índice de respuesta a la inmunoterapia. Se ha propuesto tener en cuenta la heterogeneidad de estas mutaciones para aumentar la precisión de este biomarcador (Gibney et al., 2016).
- **Repertorio de TCR:** El estudio del repertorio de TCR informa sobre la capacidad linfocitaria para el reconocimiento de neoantígenos tumorales. Aquellos clones capaces de reconocer a estos antígenos se expandirán y con ello se podrá detectar la inmunogenicidad del tumor y la respuesta a tratamientos como la inmunoterapia. El repertorio se puede caracterizar por varios parámetros obtenidos con técnicas como la secuenciación masiva, y además se pueden comparar las similitudes entre varios repertorios (distintos individuos, compartimentos o momentos) (Ruggiero et al., 2015; Schrama et al., 2017).

La caracterización del repertorio de TCR supone un biomarcador dinámico ya que permite la monitorización de los cambios, incluso de una forma poco invasiva como pueden ser los repertorios de sangre periférica.

3. NADIM

NADIM es un ensayo clínico multicéntrico cuyo objetivo era introducir una nueva opción de tratamiento para los pacientes CPCNP de estadio IIIA resecables y evaluar su capacidad antitumoral y seguridad. Este tratamiento consistiría en una neoadyuvancia de quimio e inmunoterapia administrada antes de la intervención quirúrgica para una resección completa del tumor.

Los resultados obtenidos fueron que de los 46 reclutados solo el 20% habían progresado o muerto al momento de corte de seguimiento (con una mediana de seguimiento de 24 meses) por lo que se conseguían unas buenas supervivencias tanto la global (90%) como la supervivencia libre de progresión (77%). Además, la tolerancia al tratamiento fue en general buena y los efectos secundarios no supusieron retrasos o cancelaciones de la cirugía.

Los pacientes que, además consiguieron una resección o una respuesta patológica completa (CPR, *Complete Pathological Response*), con un 0% de células tumorales viables en el tumor primario y ganglios analizados, presentaban una mejor supervivencia libre de progresión que aquellos pacientes con una respuesta patológica mayor (NCPR, *Non-Complete Pathological Response*), caracterizados por tener menos de un 10% de células tumorales viables; y que los pacientes con una respuesta patológica incompleta (IPR, *Incomplete Pathological Response*, >10% de células viables).

Posteriormente a estos resultados, se realizaron otros ensayos clínicos para poder validar esta información con una cohorte mayor de pacientes y comparar este tratamiento con una cohorte control recibiendo una neoadyuvancia solo con quimioterapia. Estos ensayos son el Checkmate 816 (NCT02998528) y el NADIM II (NCT03838159).

El ensayo Checkmate 816 consiguió relacionar el tratamiento neoadyuvante de quimio-inmunoterapia con una mejor supervivencia libre de eventos (o de cualquier tipo de progresión o recurrencia de la enfermedad) y un mayor porcentaje de pacientes que alcanzaban una CPR. A raíz de estos resultados, en el año 2022, la FDA (*Food and Drug Administration*) autorizó esta neoadyuvancia para pacientes CPCNP con un tumor resecable, por lo que ahora supone el estándar de tratamiento de estos pacientes (Center for Drug Evaluation and Research, 2022; Forde et al., 2022; Provencio et al., 2020).

Tratamiento Adyuvante:

En cuanto a esta fase del tratamiento, administrada tras la cirugía, todavía quedan muchas cuestiones por resolver como qué opciones se podrían proponer para mejorar este tratamiento como la duración, la administración de fármacos en combinación, ... También surge la posibilidad de personalizar este tratamiento o adaptarlo según la respuesta o situación de la enfermedad tras la neoadyuvancia ya que actualmente las decisiones terapéuticas actuales se basan en el estado basal del tumor (antes incluso del tratamiento neoadyuvante). Para llevar esto a cabo es necesario el estudio y desarrollo de nuevos biomarcadores enfocados en la adyuvancia (Provencio, Calvo, et al., 2022).

Estudio Previo del Repertorio de TCR en pacientes NADIM:

En el año 2021 se publicaron los resultados relativos a la caracterización del repertorio de TCR de los pacientes CPCNP participantes en el estudio NADIM, particularmente los repertorios de las muestras recogidas antes y después del tratamiento neoadyuvante en sangre periférica y tejido.

Este estudio inicial reveló como determinadas métricas del repertorio eran capaces de predecir e identificar con las muestras basales de tejido qué pacientes eran los que iban a conseguir CPR de los que no (NCPR, *Non-Complete Pathological Response*) que comprendía a los pacientes MPR e IRP. Estos potenciales biomarcadores analizados por curva ROC dieron mejores resultados que los biomarcadores tradicionales (expresión PDL1 y TMB) (Casarrubios et al., 2021).

HIPÓTESIS Y OBJETIVOS:

HIPÓTESIS:

El estudio del repertorio de TCR en los pacientes CPCNP participantes en el ensayo clínico NADIM supone una fuente potencial de biomarcadores predictivos y de monitorización que informan de las diferencias del estado inmune de los pacientes CPR y NCPR.

OBJETIVOS:

Objetivo Principal →

- Estudiar el papel del repertorio de TCR como fuente potencial de biomarcadores en pacientes de CPCNP tratados con quimio-inmunoterapia neoadyuvante seguido de una inmunoterapia adyuvante.

Objetivos Secundarios →

- Caracterización del repertorio de TCR en sangre periférica recogida durante la adyuvancia y determinar cambios debido al tratamiento.
- Evaluar los cambios en las métricas que describen el repertorio de TCR a lo largo de todo el tratamiento completo de NADIM (neoadyuvancia y adyuvancia).

MATERIALES Y MÉTODOS:

Diseño del Estudio y Directrices Éticas:

Las muestras empleadas para la realización de este trabajo pertenecen a los pacientes que participaron en el ensayo clínico NADIM (NCT03081689) promovido por el Grupo Español de Cáncer de Pulmón (GECP).

NADIM es un ensayo multicéntrico de fase II, en el que fueron reclutados 46 pacientes de CPCNP que se encontraban en un estadio localmente avanzado y resecable (IIIA N2). Estos pacientes recibieron el siguiente tratamiento: una quimio-inmunoterapia neoadyuvante de 3 ciclos (consistente de paclitaxel, carboplatino y nivolumab como agente inmunoterapéutico contra PD1), posteriormente los pacientes se sometieron a una cirugía para resecar completamente el tumor y, por último, recibieron inmunoterapia adyuvante con nivolumab durante 1 año.

De estos 46 pacientes se obtuvieron muestras de sangre y tejido durante varios momentos del tratamiento, representados en el siguiente esquema (Fig. 1).

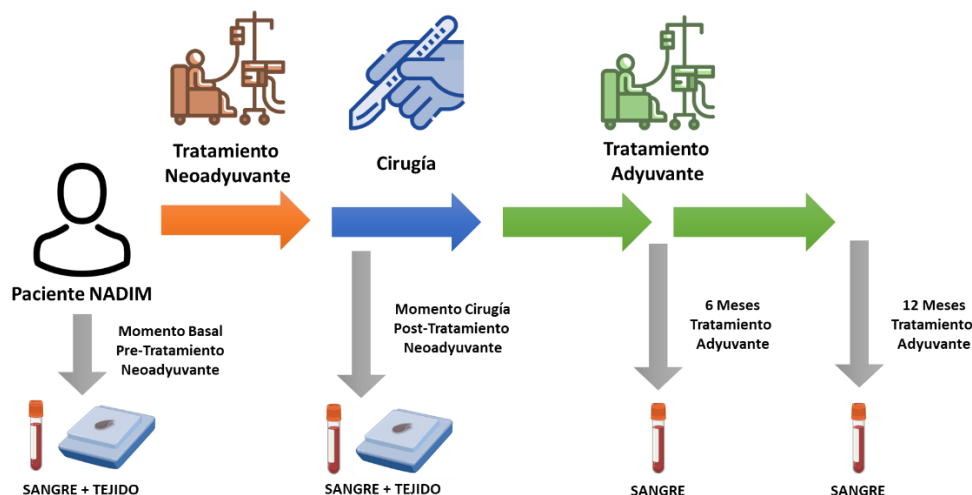


Figura 1. Esquema del tratamiento y la recogida de muestras del ensayo NADIM.

El ensayo fue realizado de acuerdo a la guías propuestas por la Conferencia Internacional de Harmonización de Buenas Prácticas Clínicas y la Declaración de Helsinki. Todos los pacientes dieron su consentimiento por escrito antes de ser reclutados. Y, además, el protocolo fue aprobado por el comité ético del Hospital Puerta de Hierro (Provencio et al., 2020).

Aislamiento de PBMCs en Sangre Periférica:

Las muestras sanguíneas fueron recogidas en tubos de EDTA de 10ml en cada uno de los hospitales incluidos para luego ser enviadas al Hospital Puerta de Hierro donde eran procesadas en un período de un máximo de 24h desde la extracción.

El proceso de obtención de las poblaciones celulares de interés (PBMCs) consistió en los siguientes pasos: dilución de la muestra 1:1 con RPMI 1640 (medio de cultivo celular), separación por gradiente de densidad con Lymphoprep (medio para gradientes de densidad, Stemcell) y una centrifugación de 1200 rpm durante 25 minutos, recogida en tubos de 15 ml, lavado con RPMI 1640, conteo con cámara de Neubauer y congelación para su conservación en nitrógeno líquido en viales de 3 a 5 millones de células con 1 ml de medio de congelación (1:1 FBS/RPMI 1640 con 10% de DMSO).

Extracción de ARN en las PBMCs:

La extracción de todas las muestras disponibles se realizó con el kit *Maxwell RSC simply RNA Cells* (Promega). Siguiendo las instrucciones del fabricante la extracción se realizó descongelando un vial de cada momento y paciente. La cuantificación de ARN obtenido se realizó con el kit *Qubit RNA BR Assay* (Thermo Fisher-Scientific) y la medición con un aparato Qubit. Las muestras obtenidas fueron almacenadas a -80°C .

Extracción de ARN en Muestras de Tejido Parafinado:

Las muestras de tejido parafinado tanto del momento del diagnóstico y la cirugía fueron enviadas desde los hospitales colaboradores del ensayo al Hospital Puerta de Hierro.

De cada una de las muestras se cortaron 2 virutas de $10\mu\text{m}$ para la extracción de ARN y ADN con el kit *truXTRAC FFPE DNA* (Covaris). Posteriormente, la cuantificación de ARN se realizó con el kit *Qubit RNA BR Assay* (Thermo Fisher-Scientific). Estas muestras también fueron almacenadas a -80°C .

Preparación de las Librerías y Secuenciación:

Este proceso sigue tres pasos: generación de librerías por AmpliSeq, carga del chip de secuenciación con las librerías por el aparato Ion Chef (Thermo Fisher-Scientific), y la secuenciación del chip cargado con el secuenciador Ion S5 System.

Las librerías de TCR obtenidas a partir del ARN de las **PBMCs**, fueron realizadas con **Oncomine TCR Beta-LR**. Este ensayo utilizó 25ng de ARN que fue transcrito previamente a ADNc por la enzima SuperScript IV VILO Master Mix (Thermo Fisher-Scientific). Este panel incluye 64 cebadores que reconocen la región CDR1 del gen variable y 2 cebadores que reconocen la región constante de la cadena β del TCR, lo que genera amplicones de 330 pares de bases correspondientes a las regiones CDR1, CDR2 y CDR3 (Fig. 2A). En cuanto a la carga del chip se utilizaron 8 librerías de 25 pmol/L y un volumen similar de cada una para una misma mezcla, de la cual se cargaron 25 μl en un chip Ion 530 utilizando el Ion Chef.

Por otro lado, las librerías de TCR de las muestras de **tejido parafinado** se realizaron con **Oncomine TCR Beta-SR**. Este ensayo utiliza un ARN inicial de 100ng, transcrito previamente a ADNc por la enzima SuperScript IV VILO Master Mix (Thermo Fisher-Scientific). A diferencia del anterior, los cebadores de este panel consisten en 59 cebadores que reconocen la región FR3 (final de la región CDR2 del gen variable) y 14 cebadores que reconocen el gen de unión, generando un amplicón de 80 pares de bases correspondientes a la región CDR3 del TCR (Fig. 2B). Para su carga en el chip, se agrupan 32 librerías (25 pmol/L) con un volumen similar en una misma mezcla, de la que se cargan 25 μl en un chip Ion 540 con el Ion Chef.

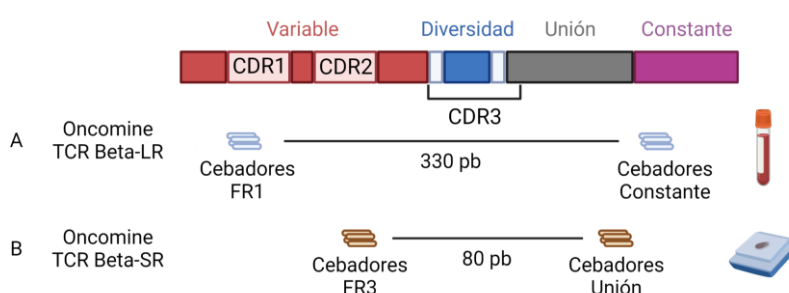


Figura 2. Esquema de cebadores y regiones secuenciadas por los paneles Oncomine TCR Beta-LR (A) y Oncomine TCR Beta-SR (B).

Análisis Bioinformático:

El procesamiento inicial de los archivos BAM procedentes de la secuenciación se realizó con la herramienta Ion Reporter en su versión 5.12 (Thermo Fisher-Scientific).

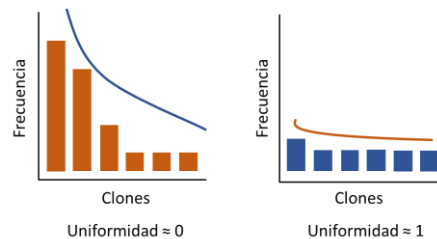
Las secuencias de las librerías de TCR se alinearon con la herramienta Ion Reporter Oncomine TCR Beta en su versión 1.2. Esta herramienta utilizaba la base de datos IMGT como referencia de los genes variables, diversidad y unión del TCR para obtener las frecuencias, las secuencias, las métricas y el número de clonotipos identificados en cada una de las muestras. Las métricas analizadas fueron:

1. **Índice de Diversidad de Shannon:** Calculada por la siguiente fórmula, donde p_i es la frecuencia del clonotipo i para la muestra con un número N de clonotipos únicos.

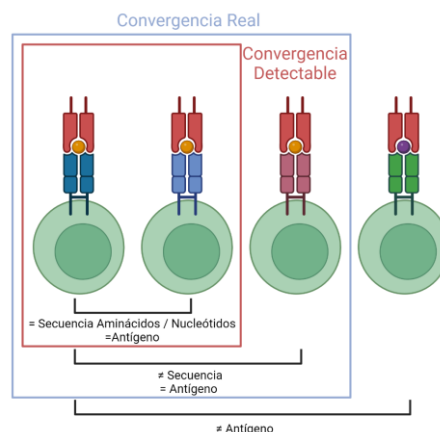
$$\text{Índice de Diversidad de Shannon} \rightarrow H = - \sum_{i=1}^N p_i \log_2(p_i)$$

2. **Uniformidad:** Describe cómo de equilibrado está distribuido el repertorio de TCR. Si este valor es muy cercano a 0 significa que el repertorio no es equilibrado debido a un pequeño número de clones muy predominantes, en cambio, si es cercano a 1, el repertorio es equilibrado y todos los clones presentan frecuencias similares. Es calculado a partir del dato anterior, siguiendo la siguiente fórmula, siendo N el número total de clones detectados en la muestra.

$$\text{Uniformidad} = \frac{H}{\log_2(N)}$$



3. **Convergencia:** Se define como la frecuencia agregada de los clones que son capaces de reconocer al mismo antígeno, pero solo es detectable en aquellos que compartan al menos la misma secuencia de aminoácidos (su secuencia de nucleótidos puede ser distinta).



La clasificación por frecuencias de los clones de cada muestra se realizó utilizando como valor de referencia su abundancia relativa en el repertorio del TCR total que fue denominado espacio clonal. De esta forma, los clones quedaron divididos por las categorías top 1%, top 1-2%, top 2-5% y top >5% (Casarrubios et al., 2021; Looney et al., 2020).

Análisis Estadístico:

Los análisis estadísticos fueron realizados a través de RStudio (R Core Team, 2022) con los paquetes básicos instalados por defecto y otros como “dplyr” para facilitar el manejo de las bases de datos (Wickham, 2022). Por otro lado, para las gráficas se utilizó el software GraphPad Prism 9.

Los test estadísticos empleados fueron principalmente el test U de Mann-Whitney para la comparación de diferencias estadísticas entre las muestras de pacientes CPR y NCPR; y el test de Wilcoxon para el estudio de la evolución de los datos a lo largo del tratamiento en las muestras pareadas disponibles.

Los scripts empleados para llevar a cabo los análisis mencionados se encontrarán en el anexo 1.

RESULTADOS:

Diferencias de los repertorios de TCR en pacientes CPR y NCPR:

Tal y como se realizó en el estudio de las muestras recogidas en la neoadyuvancia, el primer enfoque del trabajo fue analizar las métricas obtenidas de los archivos iniciales con Ion Reporter y el espacio clonal del top 1% de los repertorios de TCR en las muestras disponibles de los 46 pacientes. Se escogió estudiar el top 1% de los clones, ya que en el estudio de la neoadyuvancia se demostró su potencial como biomarcador predictivo de respuesta. En este análisis se compararon las características de estos repertorios entre los pacientes CPR y NCPR mediante el test estadístico U de Mann-Whitney.

Como ya se conocía, los resultados de la neoadyuvancia en tejido revelaron que los pacientes CPR presentaban unos repertorios menos uniformes y con un espacio clonal mayor para el top 1% de sus clones (Fig. 1A) (Casarrubios et al., 2021).

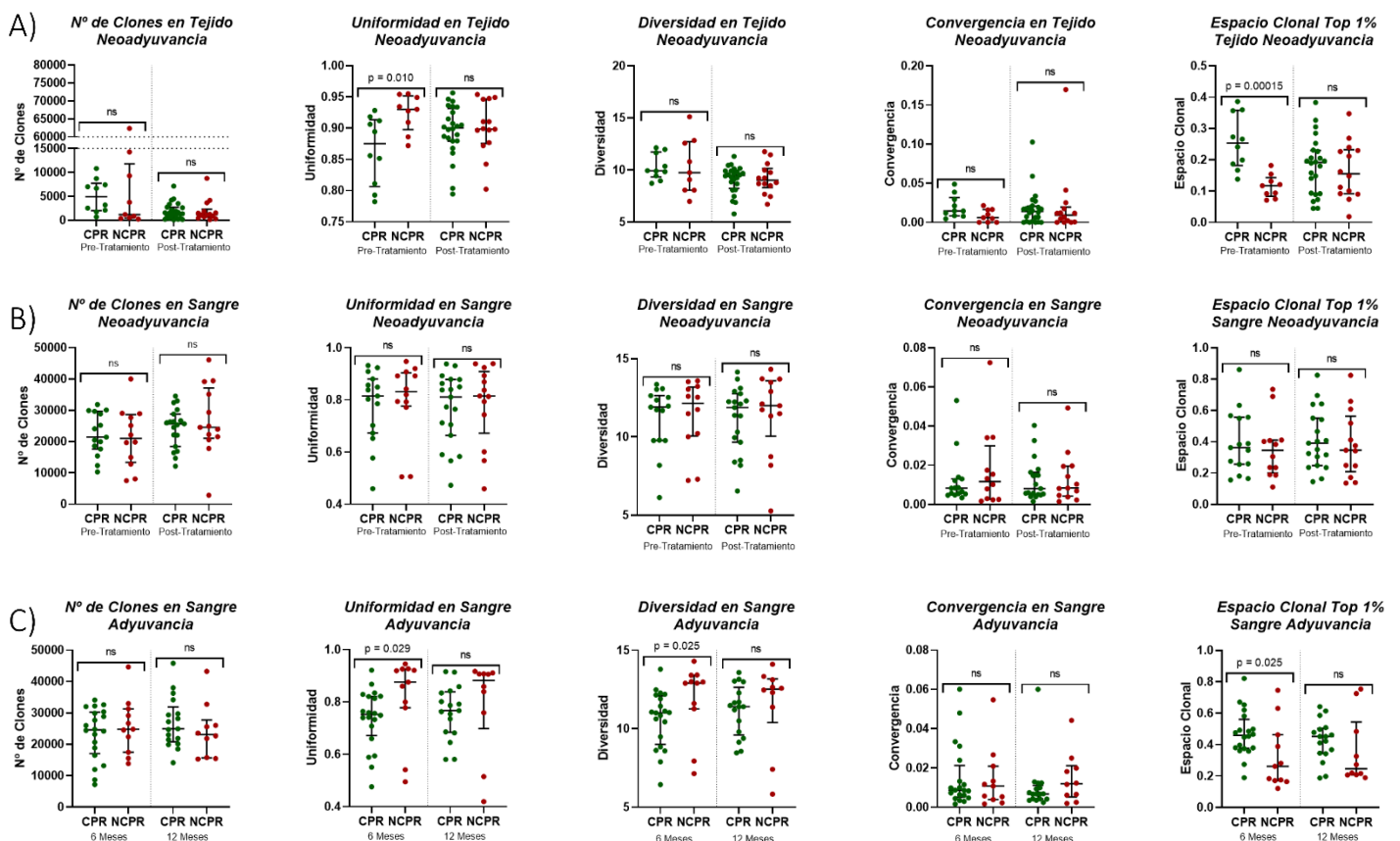


Figura 1. Caracterización del repertorio de TCR en tejido y sangre periférica según la respuesta patológica de los pacientes. A) Distribución de las métricas (número de clones, uniformidad, diversidad y convergencia) y el espacio clonal de los clones del top 1% del repertorio de TCR en el tejido antes y después del tratamiento neoadyuvante según la respuesta patológica de los pacientes. B) Distribución de las métricas y el espacio clonal del top 1% del repertorio de TCR en la sangre periférica antes y después del tratamiento neoadyuvante según la respuesta patológica de los pacientes. C) Distribución de las métricas y el espacio clonal del top 1% del repertorio de TCR en la sangre periférica a los 6 y 12 meses del inicio del tratamiento adyuvante según la respuesta patológica de los pacientes.

Al incorporar al análisis las muestras sanguíneas obtenidas durante el tratamiento adyuvante se obtuvieron nuevos resultados. En las muestras recogidas a los 6 meses de tratamiento (n = 31), los repertorios de TCR en pacientes CPR (n = 20) poseían unos repertorios sanguíneos menos uniformes (p = 0.029), menos diversos (p = 0.025) y con un mayor espacio clonal para sus clones del top 1% (p = 0.025) que los pacientes NCPR (n = 11). Para el resto de métricas no se observaron diferencias significativas entre estos dos grupos de pacientes (Fig. 1C).

En el análisis de las muestras recogidas a los 12 meses (n = 27) de tratamiento no se vieron diferencias significativas en ninguna de las métricas o en el espacio clonal ocupado por el top 1% de sus clones entre los pacientes CPR (n = 17) y NCPR (n = 10) (Fig. 1C).

Impacto de los tratamientos neoadyuvante y adyuvante en las métricas de TCR:

El siguiente paso fue el análisis de la evolución de esas métricas y espacio clonal a lo largo del tratamiento con las muestras pareadas disponibles mediante el test estadístico de Wilcoxon; primero de forma general sin estratificar a los pacientes por respuesta y posteriormente revisando por separado a los pacientes CPR y NCPR.

En los análisis del repertorio de TCR tisular, el análisis no estratificado para ver los cambios generales que se producían durante la neoadyuvancia en el repertorio de TCR tisular (n = 18 pacientes con muestras pareadas) mostró un descenso en el número de clones (p = 0.014) y en la diversidad (p = 0.043) tras la neoadyuvancia (Fig. 2A).

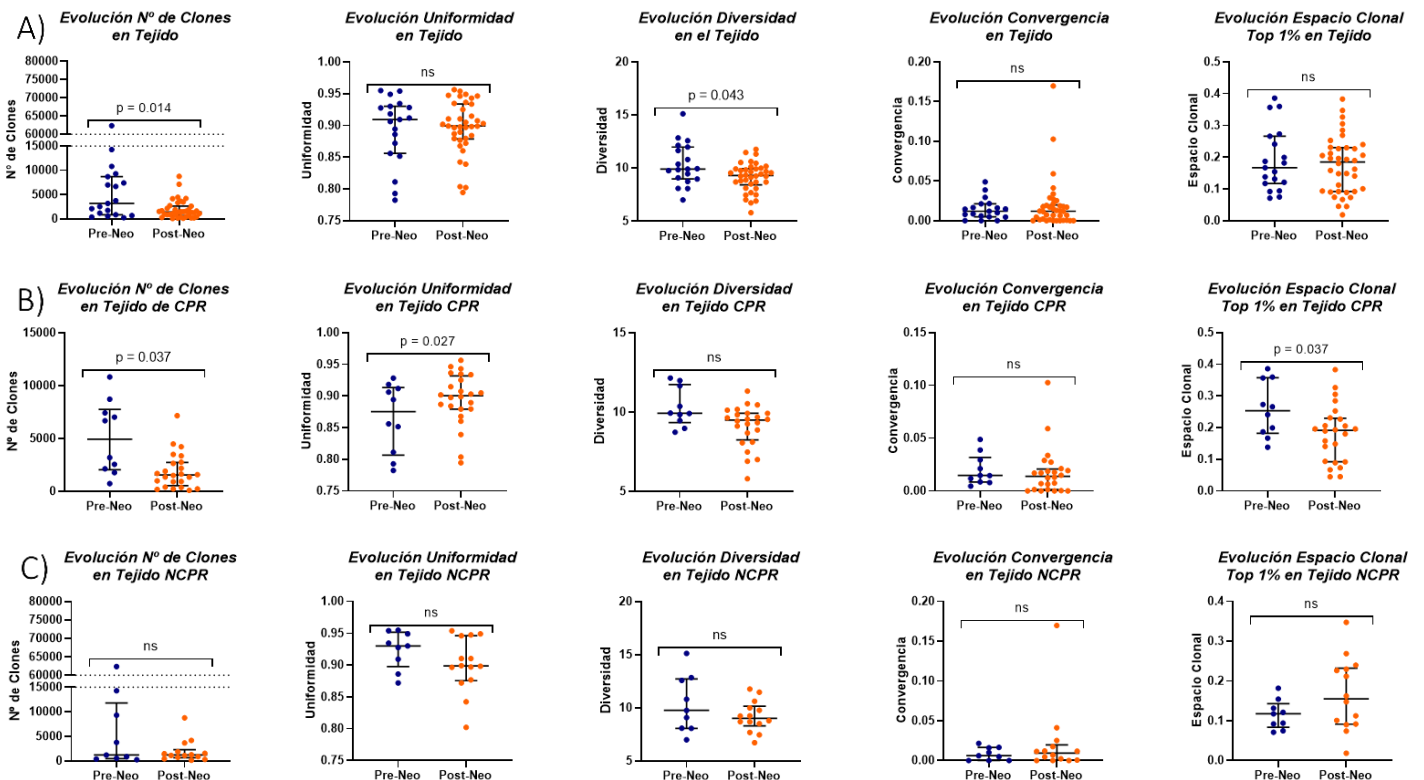


Figura 2. Evolución de las métricas y el espacio clonal del top 1% en los repertorios de TCR tisulares durante el tratamiento neoadyuvante. A) Evolución de los datos sin estratificar a los pacientes según su respuesta patológica. B) Evolución del repertorio tisular en los pacientes CPR durante la neoadyuvancia. C) Evolución del repertorio tisular en los pacientes NCPR durante la neoadyuvancia.

Por otro lado, los análisis de la evolución de los pacientes CPR ($n = 10$) revelaron descensos en el número de clones ($p = 0.037$) y en el espacio clonal del top 1% ($p = 0.037$), y un aumento en la uniformidad ($p = 0.027$) como ya se había comprobado en el estudio de la neoadyuvancia (Casarrubios et al., 2021) (Fig. 2B y 4A). Los pacientes NCPR ($n = 8$) no sufrieron diferencias significativas durante esta fase del tratamiento (Fig 2C y 4A).

En cuanto al análisis de los cambios de los repertorios sanguíneos (Fig. 3), el análisis se dividió por etapas para estudiar los efectos de cada aspecto del tratamiento (neoadyuvancia, adyuvancia, ...) y dentro de cada etapa se analizaron todos pacientes y en base a la respuesta patológica alcanzada. El primer análisis consistió en estudiar el efecto de la participación en el ensayo NADIM para los pacientes, es decir, teniendo en cuenta tanto la neoadyuvancia como la adyuvancia posterior para los pacientes (Fig. 4B), para ello se aplicó el test de Wilcoxon para comparar las muestras basales con cada de una de las siguientes muestras disponibles y pareadas.

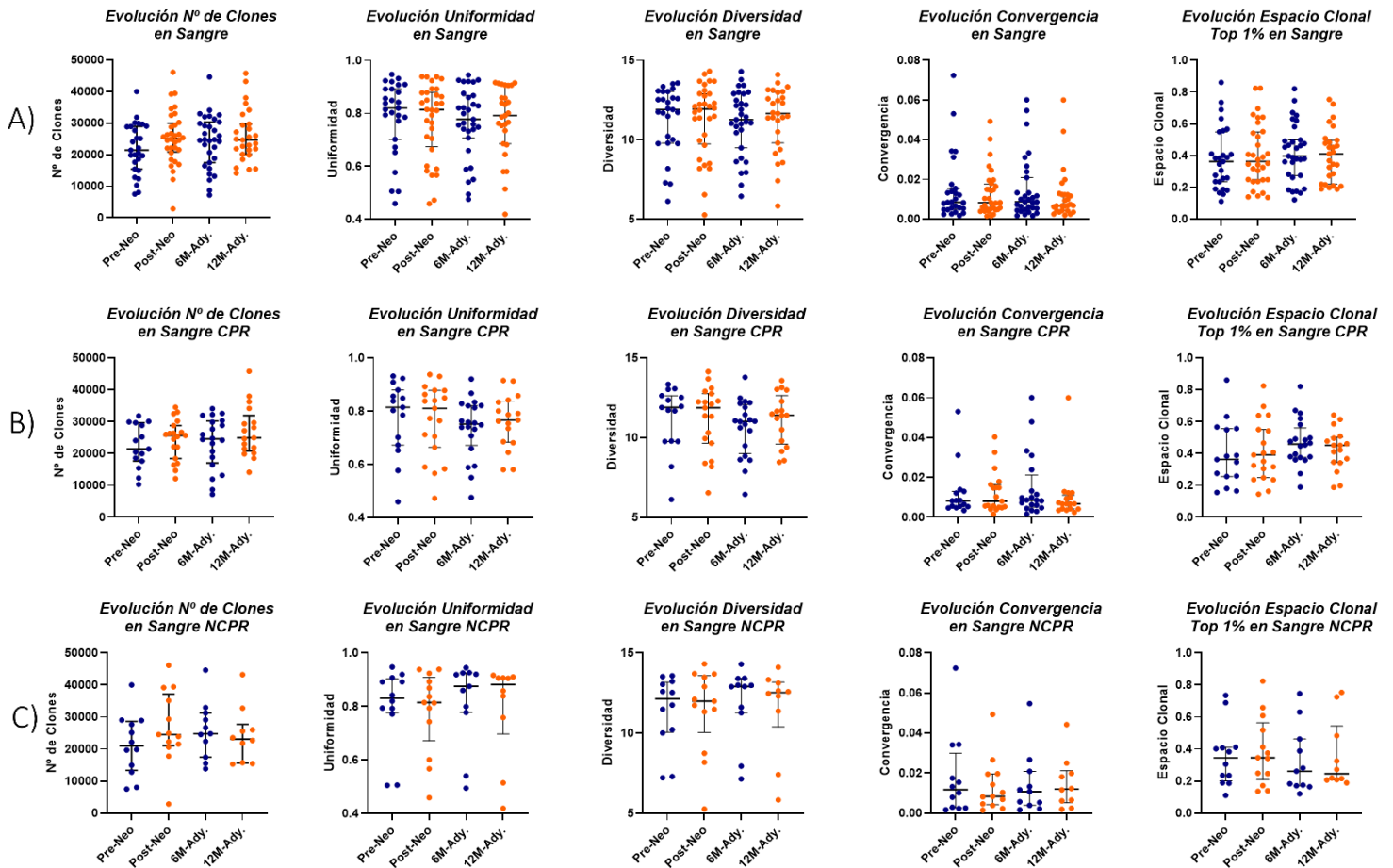


Figura 3. Evolución de las métricas y el espacio clonal del top 1% de los repertorios de TCR en sangre periférica durante el ensayo NADIM. A) Evolución de los datos sin estratificar a los pacientes según su respuesta patológica durante todo el ensayo. B) Evolución de los repertorios sanguíneos en los pacientes CPR durante todo el ensayo. C) Evolución de los repertorios sanguíneos en los pacientes NCPR durante todo el ensayo.

Los resultados de este análisis mostraron que a los 6 meses del inicio de la adyuvancia había un aumento general del número de clones ($p = 0,046$) en el análisis de todos los pacientes ($n = 21$). Separando por respuesta, en los pacientes CPR ($n = 12$) se observaba un descenso en su uniformidad ($p = 0,016$) y diversidad ($p = 0,042$), y un ascenso en el espacio clonal del top 1% de clones ($p = 0,012$); mientras tanto en los pacientes NCPR no se vieron cambios en este momento del tratamiento. En el resto de fases del tratamiento no se detectaron otros cambios significativos (Fig. 4B).

En el siguiente análisis se pretendían estudiar los efectos y cambios provocados por la adyuvancia comparando la muestra del fin de la neoadyuvancia con las muestras recogidas durante la administración del tratamiento adyuvante. Sin embargo, no se observó ningún cambio significativo en la comparación de estas muestras (Fig. 4C).

Por último, se compararon las muestras obtenidas durante la adyuvancia para estudiar los efectos de la duración del tratamiento adyuvante. Entre los 6 y 12 meses del tratamiento adyuvante los repertorios de los pacientes NCPR ($n = 10$) se observó un descenso de la uniformidad ($p = 0,037$) y de la diversidad ($p = 0,049$), mientras que el espacio clonal de su top 1% de los clones aumentó ($p = 0,020$) (Fig. 4D).

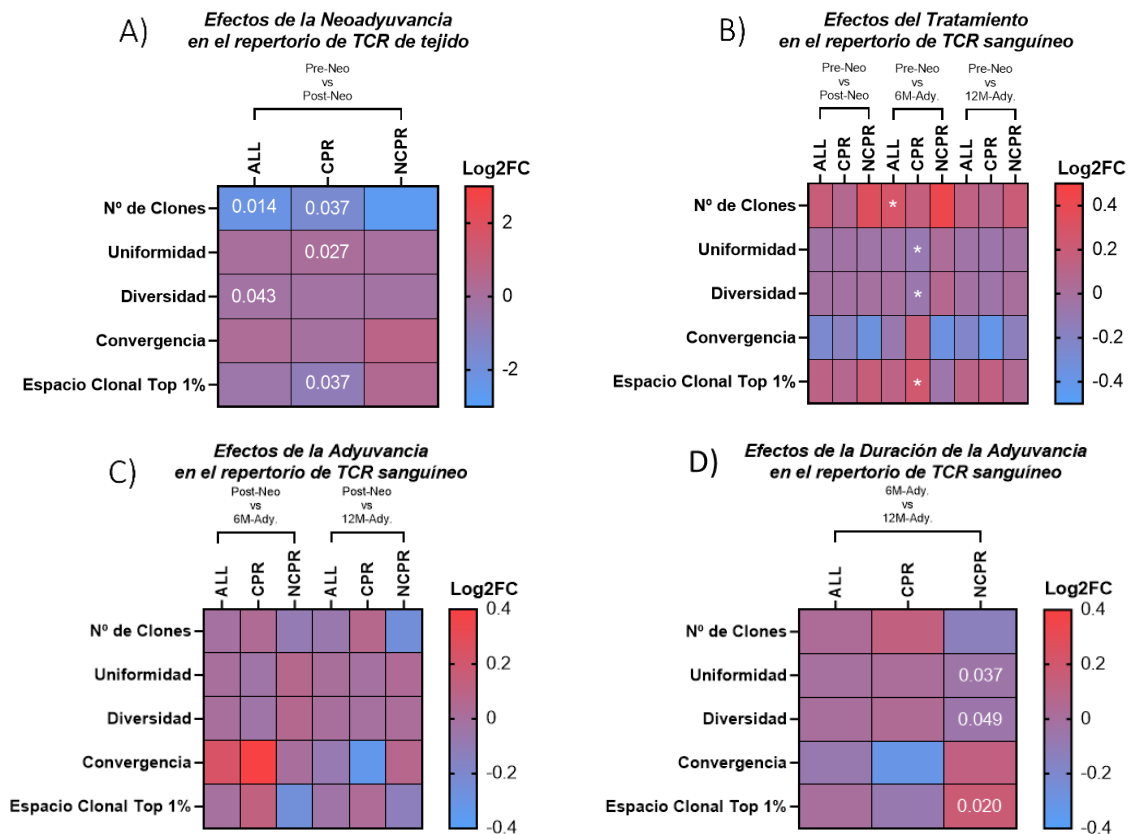


Figura 4. Heatmaps representados los cambios sufridos en los repertorios de TCR a lo largo de todo el ensayo. A) Cambios en el repertorio de tejido durante el tratamiento neoadyuvante. B) Efectos del tratamiento del ensayo NADIM en los valores basales del repertorio sanguíneo. C) Efectos del tratamiento adyuvante tras la neoadyuvancia en el repertorio sanguíneo. D) Efectos de la duración del tratamiento adyuvante en el repertorio sanguíneo.

Seguimiento de los Clones del Top 1% de la Muestra Basal de Tejido:

La siguiente parte del análisis consistió en un seguimiento del espacio clonal de los clones pertenecientes al top 1% de la muestra basal de tejido, que representarían los principales clones que se encontrarían en la zona tumoral en su estado basal antes de comenzar el tratamiento. Este análisis es posible gracias a que cada clon es identificado con los datos obtenidos con su secuenciación de TCR, determinando su identidad gracias a la suma de su secuencia nucleotídica de la zona CDR3 y los genes de la región variable de la cadena β . Este TCR es único para cada clon y no cambia en ningún momento ni para sus células hija, por lo que es posible seguir clones determinados a lo largo del tratamiento en los diferentes compartimentos.

Además, como hemos visto anteriormente el espacio clonal ocupado por estos clones predominantes de tejido tenían la capacidad de diferenciar a aquellos pacientes que llegarán a una CPR de los que no (Casarrubios et al., 2021).

Realizando un análisis previo con el test estadístico U de Mann-Whitney, se comparó el valor del espacio clonal de aquellos clones identificados como top 1% en la muestra basal tisular en los pacientes CPR y NCP. Los resultados mostraron unas tendencias en las muestras sanguíneas, desde la cirugía hasta la adyuvancia, que revelaban que los pacientes CPR tenían un repertorio donde estos clones ocupaban un espacio clonal más elevado que los pacientes NCP. Finalmente se alcanzó una diferencia significativa a los 12 meses de tratamiento adyuvante ($p = 0.048$, $n = 12$, 7 CPR y 5 NCP) (Fig. 5B).

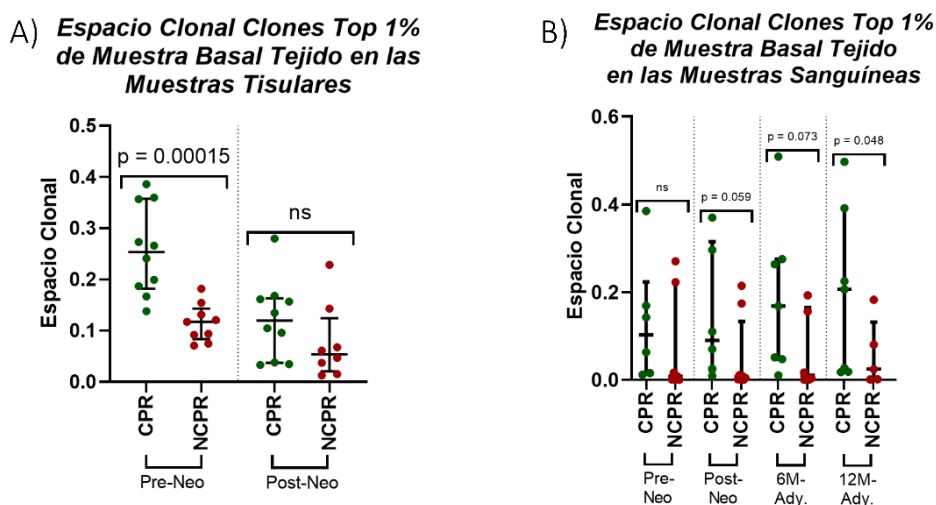


Figura 5. Espacio clonal de los clones del top 1% de la muestra basal tisular en el resto de muestras según la respuesta patológica de los pacientes. A) Espacio clonal de los clones top 1% de la muestra basal tisular durante la neoadyuvancia en las muestras de tejido según la respuesta patológica. B) Espacio clonal de los clones top 1% de la muestra basal tisular durante todo el tratamiento neoadyuvante y adyuvante en las muestras sanguíneas según la respuesta patológica.

Posteriormente se procedió al estudio de la evolución de este espacio clonal ocupado por clones pertenecientes al top 1% en el tejido al diagnóstico en las muestras de sangre a lo largo de todo el tratamiento mediante el test de Wilcoxon para las muestras pareadas disponibles. Como en los análisis anteriores, este estudio se dividió en las mismas tres etapas para evaluar los cambios y efectos de determinados momentos del tratamiento, tomando como referencia las muestras de sangre al diagnóstico, post-neoadyuvancia y a los 6 meses de adyuvancia.

Sin embargo, en ninguno de los análisis se consiguieron ver cambios significativos (Fig. 6). Aunque no es estadísticamente significativo, parece haber una tendencia a que los pacientes CPR mantengan un aumento de este espacio clonal top1% tejido basal en sangre a lo largo de todas las etapas del tratamiento, siendo este máximo a 6 meses. Mientras que en los pacientes NCPR se producen disminuciones de este espacio clonal a lo largo del tratamiento, sin embargo, sería necesario analizar estadísticamente con un test los fold change para confirmarlo.

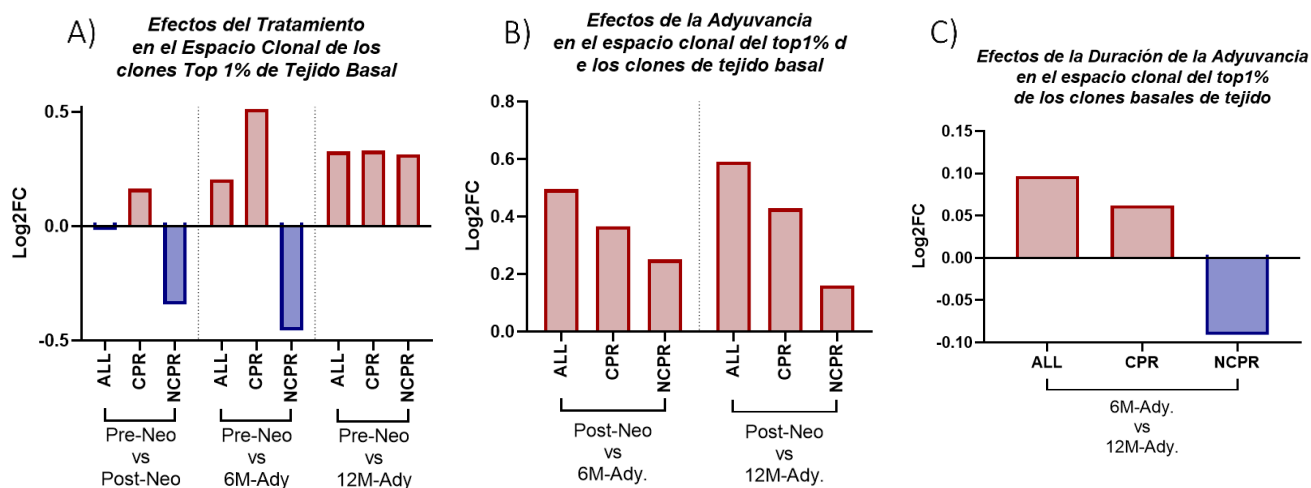


Figura 6. Evolución del espacio clonal del top 1% de los clones de la muestra basal tisular a lo largo del tratamiento en la sangre. A) Efectos del tratamiento del ensayo NADIM en el espacio clonal de las muestras sanguíneas. B) Efectos del tratamiento adyuvante tras la neoadyuvancia en el repertorio sanguíneo. C) Efectos de la duración del tratamiento adyuvante en el espacio clonal de las muestras sanguíneas.

Seguimiento de los Clones del Top 1% de la Muestra Post-Neoadyuvancia de Tejido:

Por último, se hizo un seguimiento de los clones del top 1% de la muestra de tejido obtenida en la cirugía y tras la finalización de tratamiento neoadyuvante. De esta manera estudiamos los clones predominantes en la zona tumoral tras el fin de la neoadyuvancia y vemos el efecto que tiene la adyuvancia en el espacio tumoral que ocupan.

Como en el seguimiento anterior, lo primero fue comparar los espacios clonales de este grupo de clones a lo largo del tratamiento adyuvante entre los pacientes CPR y NCPR (test U de Mann-Whitney), aunque este no mostró diferencias significativas entre los grupos a ningún momento del tratamiento (Fig. 7).

Y también, se procedió a un análisis de la evolución de estos clones a lo largo de este tratamiento; separándolo por un estudio de los efectos de la adyuvancia desde el momento de la cirugía (Fig. 8A), y otro estudio para estudiar el efecto de la duración de la adyuvancia en las muestras sanguíneas pareadas disponibles (test Wilcoxon) (Fig. 8B).

Espacio Clonal
Clones Top 1% Post-Neo. Tejido
en Sangre

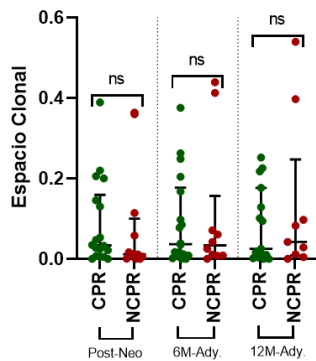


Figura 7. Espacio clonal de los clones del top 1% de la muestra posterior a la neoadyuvancia de tejido en las muestras posteriores sanguíneas según la respuesta patológica de los pacientes.

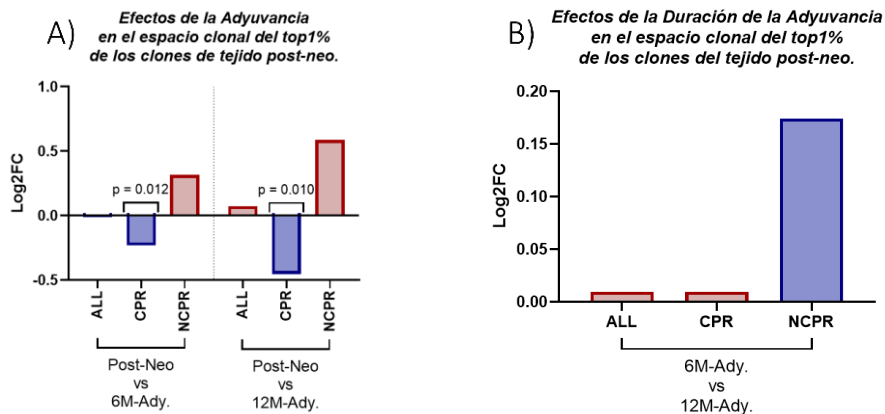


Figura 8. Evolución del espacio clonal del top 1% de los clones de la muestra tisular post-neoadyuvancia durante la adyuvancia. A) Efectos del tratamiento adyuvante tras la neoadyuvancia en el repertorio sanguíneo. C) Efectos de la duración del tratamiento adyuvante en el espacio clonal de las muestras sanguíneas.

En el estudio de los efectos y cambios producidos por el tratamiento adyuvante, se observó un descenso significativo de este espacio clonal en los pacientes CPR con respecto a la muestra sanguínea de post-neoadyuvancia tanto a los 6 meses ($n = 15$, $p = 0,012$) como a los 12 meses de tratamiento ($n = 11$, $p = 0,010$). En el resto de análisis no se detectaron otros cambios significativos.

DISCUSIÓN:

Las decisiones sobre el tratamiento adyuvante para los pacientes CPCNP con potencial operable están basadas en la misma información que se tiene en cuenta para el tratamiento neoadyuvante como lo son el estadio tumoral y el estado general del paciente.

Actualmente, ha surgido el debate de cuál podría ser la mejor manera de optimizar este tratamiento y si realmente todos los pacientes necesitan de esta terapia, ya que no se han conseguido unos resultados tan determinantes como con la neoadyuvancia. Por ello, surge la necesidad de estudiar e introducir biomarcadores predictivos que informen del estado tumoral antes del inicio del tratamiento adyuvante (tras la neoadyuvancia y cirugía) para la identificación de pacientes que tienen más riesgo a recaer y/o que se beneficiarán realmente de esta terapia o la determinación de la duración necesaria como para conseguir una eliminación completa de las posibles micrometástasis que puedan desembocar en una recaída o progresión futura en el paciente (Chaft et al., 2021; Provencio, Calvo, et al., 2022).

La respuesta patológica de los pacientes participantes NADIM tras la cirugía parece determinar diferencias en la supervivencia a largo plazo. Sin embargo, aún no está claro cuáles son los mecanismos moleculares que diferencian a los pacientes CPR de los NCPR. En el estudio de los repertorios de TCR de la neoadyuvancia (que determina el reconocimiento antigénico de los linfocitos T) se identificaron biomarcadores con potencial predictivo para anticipar cuáles pacientes alcanzarían CPR y cuáles no (Casarrubios et al., 2021; Provencio et al., 2020). En cambio, el presente trabajo está dirigido a encontrar cuáles son las diferencias de estos pacientes durante el tratamiento adyuvante y cómo es la evaluación de estos repertorios a lo largo de todo el tratamiento, lo que podría tener implicaciones en la futura personalización de las terapias adyuvantes de los pacientes con CPCNP localmente avanzado.

Diferencias en las métricas y espacio clonal del top 1%:

La caracterización de los repertorios de TCR por sus métricas y su espacio clonal ocupado por el top 1% de sus clones parece suponer una fuente importante de información que distingue a los pacientes CPR de los NCPR.

El estudio de sus posibles diferencias durante el tratamiento adyuvante también marca diferencias entre los repertorios de estos pacientes a los 6 meses de tratamiento adyuvante, lo que podría sugerir una respuesta diferencial más allá de la obtenida con la neoadyuvancia. Los pacientes CPR mostraban unos repertorios menos uniformes, menos diversos y con un espacio clonal ocupado mayor para sus clones top 1%.

La uniformidad es un parámetro que informa sobre cómo de equilibrado se encuentra repartido el espacio clonal entre los clones del repertorio. Una uniformidad baja en los pacientes CPR nos indica que este reparto no es equitativo y hay un reducido número de clones que se encuentran con una frecuencia predominante frente al resto. En esta fracción de clones podrían estar incluidos aquellos clones pertenecientes al top 1%, ya que este top representa cuáles son los clones más frecuentes del repertorio y el análisis también mostró como el espacio clonal ocupado por estos clones era significativamente mayor en estos pacientes CPR.

Por otro lado, la diversidad nos informa de cuan variado es el repertorio y cómo de parecidas son sus frecuencias. Una diversidad igual a 0 indicaría que solo existe un clon en todo el repertorio, y por el contrario cuanto mayor sea su valor indicaría la presencia de una gran variedad de clones presentes en ese repertorio. Los pacientes CPR mostraban unos repertorios sanguíneos menos diversos que los pacientes NCPR. En otros estudios, se ha visto que la disminución de la diversidad está relacionada a una mejor respuesta en cáncer de pulmón metastásico, que se explica cuando hay una selección específica de clones que se expanden y empiezan a predominar en el repertorio (Öjlert et al., 2021).

Toda esta información podría entonces relacionar los tres parámetros señalados. Por lo tanto, los repertorios de TCR en pacientes CPR parecen haber sufrido una selección de unos determinados clones que se habrían expandido, lo cual habría hecho que el repertorio se desequilibrara promoviendo a unos clones más que otros. En principio estas diferencias a nivel sistémico (sangre periférica) en la respuesta inmune de los pacientes que alcanzan CPR, podría explicar la ausencia de recurrencia de la enfermedad y la larga supervivencia en estos pacientes (Forde et al., 2022; Provencio et al., 2020; Provencio, Serna-Blasco, et al., 2022).

Sin embargo, una de las limitaciones de este estudio es que no hemos concretado a que antígenos van dirigidos estos clones, por lo que no es seguro que la selección de estos clones se deba a una estrategia de inmunovigilancia tumoral. Para esta determinación es necesario contar con una tecnología de secuenciación single-cell que en combinación de herramientas como MANAFEST que es capaz de informar sobre las dinámicas de los clones específicos de antígenos tumorales recibiendo inmunoterapia (Danilova et al., 2018; Pai & Satpathy, 2021). En algunos estudios que utilizan esta información se ha generado el debate sobre origen de los clones activos contra el tumor, si se expanden en el tejido y luego pasan a la sangre, o si al contrario proviene de un repertorio nuevo sanguíneo (Ferradini et al., 1993; Yost et al., 2019; Zheng et al., 2017).

Otra información que desconocemos es el fenotipo de esos linfocitos T que podría determinar su función o el estado de su actividad. Como sabemos los linfocitos T cumplen diversas funciones dependiendo de su clase (CD4+ o CD8+) y que además cuentan con moléculas co-estimuladoras que promueven su acción e inhibitorias que la limitan. Por lo que, aunque supiéramos cuáles de esos clones están dirigidos a un ataque contra las células tumorales no sabríamos si realmente podrían llegar a cumplirlo ya que podrían estar inhibidas por puntos de control del sistema inmune como CTLA4 (como forma de resistencia adaptativa al bloqueo de PD1).

Este impedimento podría solucionarse con un análisis de citometría de flujo de las PBMCs con un marcaje fluorescente para comprobar aquellos marcadores de interés que pudieran influir con la respuesta inmunológica. Por ejemplo, en el propio ensayo NADIM se llevó a cabo este tipo de análisis para detectar biomarcadores predictivos de respuesta en la sangre recogida durante la neoadyuvancia; por ello, sería interesante hacer lo mismo para la sangre de la adyuvancia (Laza-Briviesca et al., 2021).

Evolución de las métricas y espacio clonal del top 1% a lo largo del tratamiento:

Esta fase del análisis estaba enfocada para ver los efectos que tenía el tratamiento neoadyuvante y adyuvante en los parámetros mencionados para cada uno de los momentos recogidos.

Los cambios sufridos en el tejido tras la neoadyuvancia en los pacientes CPR no se vieron reflejados en la sangre. No fue hasta los 6 meses de tratamiento adyuvante donde se vieron cambios en la sangre de los pacientes CPR. Comparando el repertorio sanguíneo basal, el repertorio a los 6 meses de la adyuvancia muestra un descenso en la uniformidad y en la diversidad, y un aumento del espacio clonal ocupado por el top 1% de los clones.

Esto apoyaría la idea de que los pacientes CPR responden a la adyuvancia con una selección y expansión específica de ciertos clones, que no ocurre en los pacientes NCPR. Aunque seguimos teniendo las mismas limitaciones mencionadas anteriormente.

Por otro lado, estudiando el efecto de la duración de la adyuvancia comparando los repertorios sanguíneos a 6 y 12 meses, los pacientes que sufren cambios son los NCPR con unos cambios similares a los mencionados anteriormente en pacientes CPR, descensos en la uniformidad y diversidad, y aumento en el espacio clonal ocupado por el top 1% de sus clones. Esto nos dice que los pacientes NCPR tardan más tiempo en responder a la terapia adyuvante, pero finalmente adquieren un comportamiento similar a los pacientes CPR. Por ello, sería importante que los pacientes NCPR si recibieran una adyuvancia con una duración de 12 meses, pero quizás con los pacientes CPR sea suficiente con 6 meses de tratamiento.

Sin embargo, faltaría una rama control para determinar si realmente estos cambios son producidos por el tratamiento. Se podría diseñar un ensayo clínico como validación de estos resultados como lo que se hizo con el tratamiento neoadyuvante entre los ensayos NADIM y NADIM II.

Seguimiento del espacio clonal de los clones predominantes en el tejido tumoral basal:

El espacio clonal de estos clones tenía un potencial como biomarcador predictivo de respuesta en el momento basal. Por lo tanto, se propuso un análisis de seguimiento para estudiar cómo evolucionaban esos clones a lo largo del tratamiento y si había un traslado importante a la sangre.

Los pacientes CPR mostraban una tendencia a poseer un espacio clonal mayor ocupado por estos clones tras la neoadyuvancia que los pacientes NCPR, siendo finalmente una diferencia significativa a los 12 meses de la adyuvancia. Podríamos suponer que los clones seleccionados y expandidos en el tejido o en los ganglios linfáticos (dado que estos pacientes se operan y se elimina el tumor primario) poco a poco se van trasladando a la sangre hasta ser diferencialmente importantes al final del tratamiento.

Todo ello indicaría que pese a eliminar el tumor de forma completa y los ganglios linfáticos regionales, a nivel sistémico se produce un aumento de estos clones predominantes del tejido, pese a que este centro de activación, el tumor y las estructuras linfoides terciarias (Schumacher, 2022), ya no está, dando relevancia a órganos linfoides secundarios (Gardner & Ruffell, 2016).

Esto vuelve a abrir el debate sobre el origen de los clones seleccionados y expandidos como respuesta al tratamiento, ya que como se ha mencionada antes hay estudios con resultados en ambas direcciones. En nuestro estudio, teniendo en cuenta nuestras limitaciones, la dirección que se reflejaría sería una selección y expansión en el tumor que finalmente se trasladarían a los repertorios sanguíneos tras todo el tratamiento.

Aunque también es cierto que se ha demostrado en el tejido tumoral estos clones son células exhaustas que pierden su actividad por varios mecanismos como el remodelado epigenético en el contexto del bloqueo de PD1, que es el principal argumento de los que defienden el reemplazo del repertorio por parte de la sangre (Ferradini et al., 1993; Pauken et al., 2016; Yost et al., 2019; Zheng et al., 2017).

Por otro lado, en el análisis de la evolución de estos clones no se vieron cambios significativos en ninguno de los grupos de pacientes ni en ningún momento del tratamiento. Esto puede ser debido a las pocas muestras pareadas disponibles y a un posible sesgo de pérdida de las muestras de pacientes progresados (una vez progresaban no se continuaba con la recogida de las muestras). Aunque esto sugiere que estos clones predominantes no se verían afectados por el tratamiento.

Seguimiento del espacio clonal de los clones predominantes en el tejido tumoral tras la neoadyuvancia:

En este último grupo de análisis se buscaba realizar un seguimiento de los clones predominantes en la muestra tisular tras la neoadyuvancia y cirugía para comprobar los efectos de la adyuvancia en ellos.

A diferencia del anterior no se vieron diferencias entre los pacientes CPR y NCPR en cuanto a los espacios clonales ocupados por estos clones. Los clones del top 1% de esta muestra tampoco mostraba diferencias entre estos pacientes por lo que no era capaz de diferenciar entre las respuestas en ninguna de las fases del tratamiento (neoadyuvancia y adyuvancia) (Casarrubios et al., 2021).

Posteriormente haciendo un seguimiento de su evolución a lo largo de la adyuvancia si que obtuvieron resultados que indicaban un descenso de estos clones en los pacientes CPR tanto a los 6 como a los 12 meses del tratamiento. Estos cambios no se reflejaban en el estudio de la duración de la adyuvancia, por lo que se podría suponer que se trata de un descenso que ocurre a los 6 meses de tratamiento y que se mantiene en unos valores similares durante todo el tratamiento adyuvante.

Puede que al resecar el tumor los clones que se encontraban allí pierdan su importancia porque las células a las que iban dirigidas han sido extraídas, salvo en el caso de los pacientes NCPR donde se aprecia una tendencia (no significativa y pendiente de confirmar con un test estadístico aplicado a los valores de fold change) de mantenimiento y enriquecimiento de estos clones ya que todavía quedan células viables tumorales.

Por lo tanto, en el caso de los pacientes CPR estos clones pierden su valor y tienden a generar un nuevo repertorio fuera de esta zona tumoral, mientras que los NCPR tenderían a mantener estos clones para poder seguir a estas posibles micrometástasis de las células tumorales sobrevivientes. En el ensayo NADIM, se llega incluso a observar una disminución significativa de las células inmunes en la muestra tisular tras la neoadyuvancia lo que apoyaría esta teoría de dispersión de los clones restantes en la zona tumoral tras la resección (Provencio et al., 2020).

La búsqueda de estudios para comparar resultados para apoyar o contradecir lo concluido es limitado ya que es el primer estudio donde se estudia el repertorio de TCR durante el tratamiento adyuvante. Pero aporta una visión interesante sobre el origen de la preferencia de los repertorios por algunos clones y su desenvoltura a la hora de hacer una respuesta contra posibles micrometástasis.

Los clones infiltrados en el tumor, una vez realizado un reconocimiento y respuesta contra las células tumorales, presentan un estado exhausto que podría limitar su acción contra nuevos focos por ello se intenta reforzar el reemplazo del repertorio con el sanguíneo (Yost et al., 2019). Esto podría explicar por qué los pacientes CPR tienen una menor tendencia a recaer que los pacientes NCPR, ya que estos últimos no harían un reemplazo completo del repertorio tisular.

CONCLUSIONES:

1. Los pacientes CPR a los 6 meses de tratamiento adyuvante presentan un repertorio de TCR menos uniforme y diverso, con un mayor espacio clonal de sus clones pertenecientes al top 1% lo que sugiere una respuesta diferencial posterior a la neoadyuvancia.
2. La inmunoterapia neoadyuvante tiene un efecto en los repertorios de TCR tisulares, que no se reflejan en sangre hasta los 6 meses de tratamiento. Este impacto en tejido y sangre es característico de los pacientes CRP, lo que refuerza las diferencias de las respuestas entre los dos grupos de pacientes.
3. Los pacientes CPR mantienen un elevado espacio clonal ocupado por los clones predominantes en el momento basal del tejido con respecto a los NCPR, con tendencia desde el momento de la cirugía hasta alcanzar su significancia al año de tratamiento adyuvante.
4. Los clones predominantes en el tejido tras la cirugía no parecen indicar diferencias entre los pacientes CPR y NCPR, pero los pacientes NCPR parecen perderlos durante el tratamiento adyuvante.

BIBLIOGRAFÍA:

- Abbas, A. K., Lichtman, A. H., & Pillai, S. (2022). Cellular And Molecular Immunology. In *Elsevier* (10th ed.). <https://www.elsevier.com/books/cellular-and-molecular-immunology/abbas/978-0-323-75748-5>
- Calvo, V., Sierra-Rodero, B., Cruz-Bermúdez, A., & Provencio, M. (2021). Role of immunotherapy in stage IIIA non-small cell lung cancer: a narrative review. *Current Challenges in Thoracic Surgery*, 3(0), 38–38. <https://doi.org/10.21037/ccts-20-82>
- Casarrubios, M., Cruz-Bermúdez, A., Nadal, E., Insa, A., del Rosario García Campelo, M., Lázaro, M., Dómine, M., Majem, M., Rodríguez-Abreu, D., Martínez-Martí, A., de Castro-Carpeño, J., Cobo, M., López-Vivanco, G., Del Barco, E., Caro, R. B., Viñolas, N., Aranda, I. B., Viteri, S., Massuti, B., ... Provencio, M. (2021). Pretreatment tissue TCR repertoire evenness is associated with complete pathologic response in patients with NSCLC receiving neoadjuvant chemoimmunotherapy. *Clinical Cancer Research*, 27(21), 5878–5890. <https://doi.org/10.1158/1078-0432.CCR-21-1200>
- Chaft, J. E., Rimner, A., Weder, W., Azzoli, C. G., Kris, M. G., & Cascone, T. (2021). Evolution of systemic therapy for stages I–III non-metastatic non-small-cell lung cancer. *Nature Reviews Clinical Oncology*, 18(9), 547–557. <https://doi.org/10.1038/s41571-021-00501-4>
- Chen, L., & Flies, D. B. (2013). Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature Reviews Immunology*, 13(4), 227–242. <https://doi.org/10.1038/nri3405>
- Danilova, L., Anagnostou, V., Caushi, J. X., Sidhom, J. W., Guo, H., Chan, H. Y., Suri, P., Tam, A., Zhang, J., Asmar, M. El, Marrone, K. A., Naidoo, J., Brahmer, J. R., Forde, P. M., Baras, A. S., Cope, L., Velculescu, V. E., Pardoll, D. M., Housseau, F., & Smith, K. N. (2018). The mutation-associated neoantigen functional expansion of specific T cells (MANAFEST) assay: A sensitive platform for monitoring antitumor immunity. *Cancer Immunology Research*, 6(8), 888–899. <https://doi.org/10.1158/2326-6066.CIR-18-0129>
- Detterbeck, F. C. (2018). The eighth edition TNM stage classification for lung cancer: What does it mean on main street? *The Journal of Thoracic and Cardiovascular Surgery*, 155(1), 356–359. <https://doi.org/10.1016/j.jtcvs.2017.08.138>
- Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The Eighth Edition Lung Cancer Stage Classification. *Chest*, 151(1), 193–203. <https://doi.org/10.1016/J.CHEST.2016.10.010>
- Duma, N., Santana-Davila, R., & Molina, J. R. (2019). Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clinic Proceedings*, 94(8), 1623–1640. <https://doi.org/10.1016/j.mayocp.2019.01.013>
- Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., & Schreiber, R. D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. *Nature Immunology*, 3(11), 991–998. <https://doi.org/10.1038/ni1102-991>
- Ferradini, L., Mackensen, A., Genevée, C., Bosq, J., Duvillard, P., Avril, M. F., & Hercend, T. (1993). Analysis of T cell receptor variability in tumor-infiltrating lymphocytes from a human regressive melanoma: Evidence for in situ T cell clonal expansion. *Journal of Clinical Investigation*, 91(3), 1183–1190. <https://doi.org/10.1172/JCI116278>

- Forde, P. M., Spicer, J., Lu, S., Provencio, M., Mitsudomi, T., Awad, M. M., Felip, E., Broderick, S. R., Brahmer, J. R., Swanson, S. J., Kerr, K., Wang, C., Ciuleanu, T.-E., Saylor, G. B., Tanaka, F., Ito, H., Chen, K.-N., Liberman, M., Vokes, E. E., ... Girard, N. (2022). Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. *New England Journal of Medicine*, 386(21), 1973–1985. <https://doi.org/10.1056/nejmoa2202170>
- Gardner, A., & Ruffell, B. (2016). Dendritic Cells and Cancer Immunity. *Trends in Immunology*, 37(12), 855–865. <https://doi.org/10.1016/j.it.2016.09.006>
- Garon, E. B., Rizvi, N. A., Hui, R., Leighl, N., Balmanoukian, A. S., Eder, J. P., Patnaik, A., Aggarwal, C., Gubens, M., Horn, L., Carcereny, E., Ahn, M.-J., Felip, E., Lee, J.-S., Hellmann, M. D., Hamid, O., Goldman, J. W., Soria, J.-C., Dolled-Filhart, M., ... Gandhi, L. (2015). Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. *New England Journal of Medicine*, 372(21), 2018–2028. <https://doi.org/10.1056/nejmoa1501824>
- Gibney, G. T., Weiner, L. M., & Atkins, M. B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *The Lancet Oncology*, 17(12), e542–e551. [https://doi.org/10.1016/S1470-2045\(16\)30406-5](https://doi.org/10.1016/S1470-2045(16)30406-5)
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/J.CELL.2011.02.013>
- Havel, J. J., Chowell, D., & Chan, T. A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nature Reviews Cancer*, 19(3), 133–150. <https://doi.org/10.1038/s41568-019-0116-x>
- Hong, B., Heuvel, A., Prabhu, V., Zhang, S., & El-Deiry, W. (2014). Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. *Current Drug Targets*, 15(1), 80–89. <https://doi.org/10.2174/1389450114666140106101412>
- Hoy, H., Lynch, T., & Beck, M. (2019). Surgical Treatment of Lung Cancer. *Critical Care Nursing Clinics of North America*, 31(3), 303–313. <https://doi.org/10.1016/j.cnc.2019.05.002>
- Laza-Briviesca, R., Cruz-Bermúdez, A., Nadal, E., Insa, A., García-Campelo, M. del R., Huidobro, G., Dómine, M., Majem, M., Rodríguez-Abreu, D., Martínez-Martí, A., De Castro Carpeño, J., Cobo, M., López Vivanco, G., Del Barco, E., Bernabé Caro, R., Viñolas, N., Barneto Aranda, I., Viteri, S., Massuti, B., ... Provencio, M. (2021). Blood biomarkers associated to complete pathological response on NSCLC patients treated with neoadjuvant chemoimmunotherapy included in NADIM clinical trial. *Clinical and Translational Medicine*, 11(7). <https://doi.org/10.1002/ctm2.491>
- Looney, T. J., Topacio-Hall, D., Lowman, G., Conroy, J., Morrison, C., Oh, D., Fong, L., & Zhang, L. (2020). TCR Convergence in Individuals Treated With Immune Checkpoint Inhibition for Cancer. *Frontiers in Immunology*, 10(January). <https://doi.org/10.3389/fimmu.2019.02985>
- Nooreldeen, R., & Bach, H. (2021). Current and Future Development in Lung Cancer Diagnosis. *International Journal of Molecular Sciences 2021, Vol. 22, Page 8661*, 22(16), 8661. <https://doi.org/10.3390/IJMS22168661>
- O'Donnell, J. S., Teng, M. W. L., & Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. *Nature Reviews Clinical Oncology*, 16(3), 151–167. <https://doi.org/10.1038/s41571-018-0142-8>

- Öjlert, Å. K., Nebdal, D., Snapkov, I., Olsen, V., Kidman, J., Greiff, V., Chee, J., & Helland, Å. (2021). Dynamic changes in the T cell receptor repertoire during treatment with radiotherapy combined with an immune checkpoint inhibitor. *Molecular Oncology*, *15*(11), 2958–2968. <https://doi.org/10.1002/1878-0261.13082>
- Pai, J. A., & Satpathy, A. T. (2021). High-throughput and single-cell T cell receptor sequencing technologies. *Nature Methods*, *18*(8), 881–892. <https://doi.org/10.1038/s41592-021-01201-8>
- Pauken, K. E., Sammons, M. A., Odorizzi, P. M., Manne, S., Godec, J., Khan, O., Drake, A. M., Chen, Z., Sen, D. R., Kurachi, M., Barnitz, R. A., Bartman, C., Bengsch, B., Huang, A. C., Schenkel, J. M., Vahedi, G., Haining, W. N., Berger, S. L., & Wherry, E. J. (2016). Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*, *354*(6316), 1160–1165. <https://doi.org/10.1126/science.aaf2807>
- Perea, F., Bernal, M., Sánchez-Palencia, A., Carretero, J., Torres, C., Bayarri, C., Gómez-Morales, M., Garrido, F., & Ruiz-Cabello, F. (2017). The absence of HLA class I expression in non-small cell lung cancer correlates with the tumor tissue structure and the pattern of T cell infiltration. *International Journal of Cancer*, *140*(4), 888–899. <https://doi.org/10.1002/ijc.30489>
- Postmus, P. E., Kerr, K. M., Oudkerk, M., Senan, S., Waller, D. A., Vansteenkiste, J., Escriu, C., & Peters, S. (2017). Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, *28*, iv1–iv21. <https://doi.org/10.1093/annonc/mdx222>
- Provencio, M., Calvo, V., Romero, A., Spicer, J. D., & Cruz-Bermúdez, A. (2022). Treatment Sequencing in Resectable Lung Cancer: The Good and the Bad of Adjuvant Versus Neoadjuvant Therapy. *American Society of Clinical Oncology Educational Book*, *42*, 711–728. https://doi.org/10.1200/edbk_358995
- Provencio, M., Nadal, E., Insa, A., García-Campelo, M. R., Casal-Rubio, J., Dómine, M., Majem, M., Rodríguez-Abreu, D., Martínez-Martí, A., De Castro Carpeño, J., Cobo, M., López Vivanco, G., Del Barco, E., Bernabé Caro, R., Viñolas, N., Barneto Aranda, I., Viteri, S., Pereira, E., Royuela, A., ... Cruz-Bermúdez, A. (2020). Neoadjuvant chemotherapy and nivolumab in resectable non-small-cell lung cancer (NADIM): an open-label, multicentre, single-arm, phase 2 trial. *The Lancet Oncology*, *21*(11), 1413–1422. [https://doi.org/10.1016/S1470-2045\(20\)30453-8](https://doi.org/10.1016/S1470-2045(20)30453-8)
- Provencio, M., Serna-Blasco, R., Nadal, E., Insa, A., García-Campelo, M. R., Casal Rubio, J., Dómine, M., Majem, M., Rodríguez-Abreu, D., Martínez-Martí, A., De Castro Carpeño, J., Cobo, M., López Vivanco, G., Del Barco, E., Bernabé Caro, R., Viñolas, N., Barneto Aranda, I., Viteri, S., Pereira, E., ... Romero, A. (2022). Overall Survival and Biomarker Analysis of Neoadjuvant Nivolumab Plus Chemotherapy in Operable Stage IIIA Non-Small-Cell Lung Cancer (NADIM phase II trial). *Journal of Clinical Oncology*, *71*(25). <https://doi.org/10.1200/JCO.21.02660>
- Riley, R. S., June, C. H., Langer, R., & Mitchell, M. J. (2019). Delivery technologies for cancer immunotherapy. *Nature Reviews Drug Discovery*, *18*(3), 175–196. <https://doi.org/10.1038/s41573-018-0006-z>

- Rosenthal, R., Cadieux, E. L., Salgado, R., Al-Bakir, M., Moore, D. A., Hiley, C. T., Lund, T., Tanić, M., Reading, J. L., Joshi, K., Henry, J. Y., Ghorani, E., Wilson, G. A., Birkbak, N. J., Jamal-Hanjani, M., Veeriah, S., Szallasi, Z., Loi, S., Hellmann, M. D., ... Kidd, A. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature*, *567*(7749), 479–485. <https://doi.org/10.1038/s41586-019-1032-7>
- Ruggiero, E., Nicolay, J. P., Fronza, R., Arens, A., Paruzynski, A., Nowrouzi, A., Ürenden, G., Lulay, C., Schneider, S., Goerdts, S., Glimm, H., Krammer, P. H., Schmidt, M., & Von Kalle, C. (2015). High-resolution analysis of the human T-cell receptor repertoire. *Nature Communications*, *6*, 1–7. <https://doi.org/10.1038/ncomms9081>
- Schrama, D., Ritter, C., & Becker, J. C. (2017). T cell receptor repertoire usage in cancer as a surrogate marker for immune responses. *Seminars in Immunopathology*, *39*(3), 255–268. <https://doi.org/10.1007/s00281-016-0614-9>
- Schumacher, T. N., & Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science*, *348*(6230), 69–74. <https://doi.org/10.1126/science.aaa4971>
- Sharpe, A. H., & Pauken, K. E. (2018). The diverse functions of the PD1 inhibitory pathway. *Nature Reviews Immunology*, *18*(3), 153–167. <https://doi.org/10.1038/nri.2017.108>
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(1), 7–33. <https://doi.org/10.3322/caac.21708>
- Tokito, T., Azuma, K., Kawahara, A., Ishii, H., Yamada, K., Matsuo, N., Kinoshita, T., Mizukami, N., Ono, H., Kage, M., & Hoshino, T. (2016). Predictive relevance of PD-L1 expression combined with CD8+ TIL density in stage III non-small cell lung cancer patients receiving concurrent chemoradiotherapy. *European Journal of Cancer*, *55*, 7–14. <https://doi.org/10.1016/j.ejca.2015.11.020>
- Trowsdale, J. (2005). HLA genomics in the third millennium. *Current Opinion in Immunology*, *17*(5), 498–504. <https://doi.org/10.1016/j.coi.2005.07.015>
- Turner, S. J., Doherty, P. C., McCluskey, J., & Rossjohn, J. (2006). Structural determinants of T-cell receptor bias in immunity. *Nature Reviews Immunology*, *6*(12), 883–894. <https://doi.org/10.1038/nri1977>
- Yost, K. E., Satpathy, A. T., Wells, D. K., Qi, Y., Wang, C., Kageyama, R., McNamara, K. L., Granja, J. M., Sarin, K. Y., Brown, R. A., Gupta, R. K., Curtis, C., Bucktrout, S. L., Davis, M. M., Chang, A. L. S., & Chang, H. Y. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nature Medicine*, *25*(8), 1251–1259. <https://doi.org/10.1038/s41591-019-0522-3>
- Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J. Y., Zhang, Q., Liu, Z., Dong, M., Hu, X., Ouyang, W., Peng, J., & Zhang, Z. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell*, *169*(7), 1342–1356.e16. <https://doi.org/10.1016/j.cell.2017.05.035>

Anexos:

Anexo 1. Scripts empleados en el análisis:

- **Extracción de las métricas de los archivos de tejido obtenidos de Ion Reporter:**

```
#Extracción de los datos de las métricas de los archivos de FFPE
```

```
#ARCHIVOS PRE FFPE
```

```
#Generación de la tabla donde meto los datos
```

```
Metricas_FFPE_PRE = data.frame(matrix(nrow = 1, ncol = 1))
```

```
colnames(Metricas_FFPE_PRE) = c("Patient")
```

```
Metricas_FFPE_PRE$Extraction = c(NA)
```

```
Metricas_FFPE_PRE$Tumor = c(NA)
```

```
Metricas_FFPE_PRE$Response = c(NA)
```

```
Metricas_FFPE_PRE$Complete_Response = c(NA)
```

```
Metricas_FFPE_PRE$Clones = c(NA)
```

```
Metricas_FFPE_PRE$Evenness = c(NA)
```

```
Metricas_FFPE_PRE$Diversity = c(NA)
```

```
Metricas_FFPE_PRE$Convergence = c(NA)
```

```
#Lectura de los archivos pre y saltar pacientes sin muestra
```

```
for (n in 1:52) {
```

```
Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras  
FFPE/Metricas/"
```

```
Ruta_Pre_1 = "-1.metrics.csv"
```

```
Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
```

```
if(!file.exists(Ruta_Pre_2)) {next}
```

```
Pre_FFPE=read.csv(Ruta_Pre_2, header = TRUE)
```

```
#Introducción de los datos
```

```
Metricas_FFPE_PRE[n,1] = n
```

```
Metricas_FFPE_PRE[n,2] = 1
```

```
Metricas_FFPE_PRE[n,6] = Pre_FFPE$Clones
```

```
Metricas_FFPE_PRE[n,7] = Pre_FFPE$Evenness
```

```
Metricas_FFPE_PRE[n,8] = Pre_FFPE$Shannon_diversity
```



```

Metricas_FFPE_PRE[n,9] = Pre_FFPE$convergent_TCR_frequency

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")

Metricas_FFPE_PRE = filter(Metricas_FFPE_PRE,
rowSums(is.na(Metricas_FFPE_PRE)) != ncol(Metricas_FFPE_PRE))

#ARCHIVOS POST FFPE

#Generación de la tabla donde meto los datos
Metricas_FFPE_POST = data.frame(matrix(nrow = 1, ncol = 1))
colnames(Metricas_FFPE_POST) = c("Patient")
Metricas_FFPE_POST$Extraction = c(NA)
Metricas_FFPE_POST$Tumor = c(NA)
Metricas_FFPE_POST$Response = c(NA)
Metricas_FFPE_POST$Complete_Response = c(NA)
Metricas_FFPE_POST$Clones = c(NA)
Metricas_FFPE_POST$Evenness = c(NA)
Metricas_FFPE_POST$Diversity = c(NA)
Metricas_FFPE_POST$Convergence = c(NA)

#Lectura de los archivos POST y salto de los pacientes sin muestra
for (n in 1:52) {
Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Metricas/"
Ruta_POST_1 = "-2.metrics.csv"
Ruta_POST_2 = paste(Ruta_General, n, Ruta_POST_1, sep = "")
if(!file.exists(Ruta_POST_2)) {next}
POST_FFPE=read.csv(Ruta_POST_2, header = TRUE)

#Introducción de los datos
Metricas_FFPE_POST[n,1] = n
Metricas_FFPE_POST[n,2] = 2

```

```

Metricas_FFPE_POST[n,6] = POST_FFPE$Clones
Metricas_FFPE_POST[n,7] = POST_FFPE$Evenness
Metricas_FFPE_POST[n,8] = POST_FFPE$Shannon_diversity
Metricas_FFPE_POST[n,9] = POST_FFPE$convergent_TCR_frequency

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")

Metricas_FFPE_POST = filter(Metricas_FFPE_POST,
rowSums(is.na(Metricas_FFPE_POST)) != ncol(Metricas_FFPE_POST))

#Introduzco datos sobre ganglio o tumor pre
Ganglio_Pre = c (7, 13, 22, 30, 38, 45, 52)
Metricas_FFPE_PRE$Tumor = 1
Metricas_FFPE_PRE[Metricas_FFPE_PRE$Patient %in% Ganglio_Pre, 3] = 0

#Introduzco datos sobre ganglio o tumor post
Ganglio_Post = c (21, 26, 29, 30, 49)
Metricas_FFPE_POST$Tumor = 1
Metricas_FFPE_POST[Metricas_FFPE_POST$Patient %in% Ganglio_Post, 3] = 0

#Combino ambas tablas
METRICAS_FFPE <- rbind(Metricas_FFPE_PRE,Metricas_FFPE_POST)
METRICAS_FFPE=METRICAS_FFPE[order(METRICAS_FFPE[,1]),]
rownames(METRICAS_FFPE) <- 1:nrow(METRICAS_FFPE)

#Introduzco datos de respuesta
Sin_Respuesta = c(1,18, 31, 44, 52)
METRICAS_FFPE[METRICAS_FFPE$Patient %in% Sin_Respuesta, 4] = 0
IPR = c(6, 12, 21, 27, 29, 43, 45)
METRICAS_FFPE[METRICAS_FFPE$Patient %in% IPR, 4] = 1
MPR = c(23, 28, 30, 33, 37, 39, 49, 50)
METRICAS_FFPE[METRICAS_FFPE$Patient %in% MPR, 4] = 2
CPR = c(2, 4, 7, 9, 11, 13, 14, 15, 16, 17, 19,
20, 22, 24, 26, 32, 34, 35, 36, 38, 40, 42,
46, 47, 48, 51)
METRICAS_FFPE[METRICAS_FFPE$Patient %in% CPR, 4] = 3

```

```
METRICAS_FFPE[METRICAS_FFPE$Patient %in% CPR, 5] = 1
```

```
METRICAS_FFPE[METRICAS_FFPE$Patient %in% MPR, 5] = 0
```

```
METRICAS_FFPE[METRICAS_FFPE$Patient %in% IPR, 5] = 0
```

```
#Genero el archivo con todos los datos
```

```
write.table(METRICAS_FFPE, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento  
TCR/Resultados/FFPE/Metrics/Metricas_FFPE.csv", sep = "\t", row.names = F)
```

- **Extracción de las métricas de los archivos de sangre obtenidos de Ion Reporter**

```
#Extracción de los datos de las métricas de los archivos de PBMCS
```

```
#Generación de la tabla donde meto los datos PRE
```

```
Metricas_PBMCS_PRE = data.frame(matrix(nrow = 1, ncol = 1))
```

```
colnames(Metricas_PBMCS_PRE) = c("Patient")
```

```
Metricas_PBMCS_PRE$Extraction = c(NA)
```

```
Metricas_PBMCS_PRE$Response = c(NA)
```

```
Metricas_PBMCS_PRE$Complete_Response = c(NA)
```

```
Metricas_PBMCS_PRE$Clones = c(NA)
```

```
Metricas_PBMCS_PRE$Evenness = c(NA)
```

```
Metricas_PBMCS_PRE$Diversity = c(NA)
```

```
Metricas_PBMCS_PRE$Convergence = c(NA)
```

```
#Lectura de los archivos pre
```

```
for (n in 1:52) {
```

```
  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras  
  PBMCS/Metricas/"
```

```
  Ruta_Pre_1 = "-1.metrics.csv"
```

```
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
```

```
  if(!file.exists(Ruta_Pre_2)) {
```

```
    next
```

```
  }
```

```
  Pre_PBMCS=read.csv(Ruta_Pre_2, header = TRUE)
```

```

#Introducción de los datos
Metricas_PBMCS_PRE[n,1] = n
Metricas_PBMCS_PRE[n,2] = 1
Metricas_PBMCS_PRE[n,5] = Pre_PBMCS$Clones
Metricas_PBMCS_PRE[n,6] = Pre_PBMCS$Evenness
Metricas_PBMCS_PRE[n,7] = Pre_PBMCS$Shannon_diversity
Metricas_PBMCS_PRE[n,8] = Pre_PBMCS$convergent_TCR_frequency

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")
Metricas_PBMCS_PRE = filter(Metricas_PBMCS_PRE,
  rowSums(is.na(Metricas_PBMCS_PRE)) != ncol(Metricas_PBMCS_PRE))

#Generación de la tabla donde meto los datos POST
Metricas_PBMCS_POST = data.frame(matrix(nrow = 1, ncol = 1))
colnames(Metricas_PBMCS_POST) = c("Patient")
Metricas_PBMCS_POST$Extraction = c(NA)
Metricas_PBMCS_POST$Response = c(NA)
Metricas_PBMCS_POST$Complete_Response = c(NA)
Metricas_PBMCS_POST$Clones = c(NA)
Metricas_PBMCS_POST$Evenness = c(NA)
Metricas_PBMCS_POST$Diversity = c(NA)
Metricas_PBMCS_POST$Convergence = c(NA)

#Lectura de los archivos POST
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
  PBMCS/Metricas/"

  Ruta_POST_1 = "-2.metrics.csv"

```

```

Ruta_POST_2 = paste(Ruta_General, n, Ruta_POST_1, sep = "")
if(!file.exists(Ruta_POST_2)) {
  next
}

```

```

POST_PBMCS=read.csv(Ruta_POST_2, header = TRUE)

```

```

#Introducción de los datos

```

```

Metricas_PBMCS_POST[n,1] = n

```

```

Metricas_PBMCS_POST[n,2] = 2

```

```

Metricas_PBMCS_POST[n,5] = POST_PBMCS$Clones

```

```

Metricas_PBMCS_POST[n,6] = POST_PBMCS$Evenness

```

```

Metricas_PBMCS_POST[n,7] = POST_PBMCS$Shannon_diversity

```

```

Metricas_PBMCS_POST[n,8] = POST_PBMCS$convergent_TCR_frequency

```

```

}

```

```

#Quito los huecos de los pacientes sin muestra

```

```

library("dplyr")

```

```

Metricas_PBMCS_POST = filter(Metricas_PBMCS_POST,
  rowSums(is.na(Metricas_PBMCS_POST)) != ncol(Metricas_PBMCS_POST))

```

```

#Generación de la tabla donde meto los datos 6M

```

```

Metricas_PBMCS_6M = data.frame(matrix(nrow = 1, ncol = 1))

```

```

colnames(Metricas_PBMCS_6M) = c("Patient")

```

```

Metricas_PBMCS_6M$Extraction = c(NA)

```

```

Metricas_PBMCS_6M$Response = c(NA)

```

```

Metricas_PBMCS_6M$Complete_Response = c(NA)

```

```

Metricas_PBMCS_6M$Clones = c(NA)

```

```

Metricas_PBMCS_6M$Evenness = c(NA)

```

```

Metricas_PBMCS_6M$Diversity = c(NA)

```

```

Metricas_PBMCS_6M$Convergence = c(NA)

#Lectura de los archivos POST
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
    PBMCS/Metricas/"

  Ruta_6M_1 = "-3.metrics.csv"
  Ruta_6M_2 = paste(Ruta_General, n, Ruta_6M_1, sep = "")
  if(!file.exists(Ruta_6M_2)) {
    next
  }

  PBMCS_6M=read.csv(Ruta_6M_2, header = TRUE)

  #Introducción de los datos
  Metricas_PBMCS_6M[n,1] = n
  Metricas_PBMCS_6M[n,2] = 3
  Metricas_PBMCS_6M[n,5] = PBMCS_6M$Clones
  Metricas_PBMCS_6M[n,6] = PBMCS_6M$Evenness
  Metricas_PBMCS_6M[n,7] = PBMCS_6M$Shannon_diversity
  Metricas_PBMCS_6M[n,8] = PBMCS_6M$convergent_TCR_frequency

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")
Metricas_PBMCS_6M = filter(Metricas_PBMCS_6M,
  rowSums(is.na(Metricas_PBMCS_6M)) != ncol(Metricas_PBMCS_6M))

#Generación de la tabla donde meto los datos 12M

```

```

Metricas_PBMCS_12M = data.frame(matrix(nrow = 1, ncol = 1))
colnames(Metricas_PBMCS_12M) = c("Patient")
Metricas_PBMCS_12M$Extraction = c(NA)
Metricas_PBMCS_12M$Response = c(NA)
Metricas_PBMCS_12M$Complete_Response = c(NA)
Metricas_PBMCS_12M$Clones = c(NA)
Metricas_PBMCS_12M$Evenness = c(NA)
Metricas_PBMCS_12M$Diversity = c(NA)
Metricas_PBMCS_12M$Convergence = c(NA)

#Lectura de los archivos POST
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
    PBMCS/Metricas/"
  Ruta_12M_1 = "-4.metrics.csv"
  Ruta_12M_2 = paste(Ruta_General, n, Ruta_12M_1, sep = "")
  if(!file.exists(Ruta_12M_2)) {
    next
  }

  PBMCS_12M=read.csv(Ruta_12M_2, header = TRUE)

  #Introducción de los datos
  Metricas_PBMCS_12M[n,1] = n
  Metricas_PBMCS_12M[n,2] = 4
  Metricas_PBMCS_12M[n,5] = PBMCS_12M$Clones
  Metricas_PBMCS_12M[n,6] = PBMCS_12M$Evenness
  Metricas_PBMCS_12M[n,7] = PBMCS_12M$Shannon_diversity
  Metricas_PBMCS_12M[n,8] = PBMCS_12M$convergent_TCR_frequency

}

```

```

#Quito los huecos de los pacientes sin muestra
library("dplyr")

Metricas_PBMCS_12M = filter(Metricas_PBMCS_12M,
  rowSums(is.na(Metricas_PBMCS_12M)) != ncol(Metricas_PBMCS_12M))

#Combino las tablas

METRICAS_PBMCS <-
  rbind(Metricas_PBMCS_PRE, Metricas_PBMCS_POST, Metricas_PBMCS_6M, Met
    ricas_PBMCS_12M)

METRICAS_PBMCS = METRICAS_PBMCS[order(METRICAS_PBMCS[,1]),]
rownames(METRICAS_PBMCS) <- 1:nrow(METRICAS_PBMCS)

#Meto los datos de respuesta

Sin_Respuesta = c(1, 18, 31, 44, 52)
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% Sin_Respuesta, 3] = 0

IPR = c(6, 12, 21, 27, 29, 43, 45)
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% IPR, 3] = 1

MPR = c(23, 28, 30, 33, 37, 39, 49, 50)
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% MPR, 3] = 2

CPR = c(2, 4, 7, 9, 11, 13, 14, 15, 16, 17, 19,
  20, 22, 24, 26, 32, 34, 35, 36, 38, 40, 42,
  46, 47, 48, 51)
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% CPR, 3] = 3

METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% CPR, 4] = 1
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% MPR, 4] = 0
METRICAS_PBMCS[METRICAS_PBMCS$Patient %in% IPR, 4] = 0

#Genero el archivo con todos los datos

write.table(METRICAS_PBMCS, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
  TCR/Resultados/PBMCS/Metrics/Metricas_PBMCS.csv", sep = "\t", row.names =
  F)

```


- **Script para la obtención de los identificadores de clones y sus frecuencias en tejido:**

```
#Extracción de los clones identificados (variable + nucleotidos) y su frecuencia

#ARCHIVOS FFPE PRE
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clone_Summary/"

  Ruta_Pre_1 = "-1.clone_summary.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

  Pre_FFPE_CS=read.csv(Ruta_Pre_2, header = TRUE)

  #Creación de la tabla con las filas necesarias
  rows = nrow(Pre_FFPE_CS)
  Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))
  colnames(Clone_Summary) = c("ID_Clon")
  Clone_Summary$Frequency = c(NA)

  #Introducción de los datos
  Clone_Summary$ID_Clon = paste(Pre_FFPE_CS$Variable,
Pre_FFPE_CS$CDR3.NT, sep = " ")
  Clone_Summary$Frequency = Pre_FFPE_CS$Frequency

  #Guardo la tabla generada
  Archivo = paste(n, "-1", ".clone_summary_analisis.csv", sep = "")
  Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clone_Summary_Analisis/", Archivo, sep = "")
  write.csv(Clone_Summary, Archivo_2, row.names = FALSE )
}
```

```
}
```

```
#ARCHIVOS FFPE POST
```

```
#Lectura de los archivos y generación de la tabla con los datos a recoger
```

```
for (n in 1:52) {
```

```
  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras  
  FFPE/Clone_Summary/"
```

```
  Ruta_Post_1 = "-2.clone_summary.csv"
```

```
  Ruta_Post_2 = paste(Ruta_General, n, Ruta_Post_1, sep = "")
```

```
  if(!file.exists(Ruta_Post_2)) {next}
```

```
  Post_FFPE_CS=read.csv(Ruta_Post_2, header = TRUE)
```

```
#Creación de la tabla con las filas necesarias
```

```
rows = nrow(Post_FFPE_CS)
```

```
Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))
```

```
colnames(Clone_Summary) = c("ID_Clon")
```

```
Clone_Summary$Frequency = c(NA)
```

```
#Introducción de los datos
```

```
Clone_Summary$ID_Clon = paste(Post_FFPE_CS$Variable,  
Post_FFPE_CS$CDR3.NT, sep = " ")
```

```
Clone_Summary$Frequency = Post_FFPE_CS$Frequency
```

```
#Guardo la tabla generada
```

```
Archivo = paste(n, "-2", ".clone_summary_analisis.csv", sep = "")
```

```
Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras  
FFPE/Clone_Summary_Analisis/", Archivo, sep = "")
```

```
write.csv(Clone_Summary, Archivo_2, row.names = FALSE )
```

```
}
```

- **Script para la selección de los clones del top 1% en las muestras de tejido:**

```

#Clones Top 1%

#ARCHIVOS FFPE PRE
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clone_Summary_Analisis/"

  Ruta_Pre_1 = "-1.clone_summary_analisis.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

  #Calculo el número de clones del top 1 y genero una nueva tabla solo con ellos
  Pre_FFPE_CS=read.csv(Ruta_Pre_2, header = TRUE)
  Pre_FFPE_CS=Pre_FFPE_CS[order(Pre_FFPE_CS[,2], decreasing = TRUE ),]
  rownames(Pre_FFPE_CS) <- 1:nrow(Pre_FFPE_CS)
  Top1 = nrow(Pre_FFPE_CS) / 100
  Top1 = round(Top1)
  Top1_Pre_FFPE = head(Pre_FFPE_CS, Top1)

  #Guardo la tabla generada
  Archivo = paste(n, "-1", ".top1.csv", sep = "")
  Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/", Archivo, sep = "")
  write.csv(Top1_Pre_FFPE, Archivo_2, row.names = FALSE )

}

#ARCHIVOS FFPE POST
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

```

```

Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clone_Summary_Analisis/"

Ruta_Post_1 = "-2.clone_summary_analysis.csv"

Ruta_Post_2 = paste(Ruta_General, n, Ruta_Post_1, sep = "")

if(!file.exists(Ruta_Post_2)) {next}

#Calculo el número de clones del top 1 y genero otra tabla solo con ellos
Post_FFPE_CS=read.csv(Ruta_Post_2, header = TRUE)
Post_FFPE_CS=Post_FFPE_CS[order(Post_FFPE_CS[,2], decreasing = TRUE ),]
rownames(Post_FFPE_CS) <- 1:nrow(Post_FFPE_CS)

Top1 = nrow(Post_FFPE_CS) / 100
Top1 = round(Top1)
Top1_Post_FFPE = head(Post_FFPE_CS, Top1)

#Guardo la tabla generada

Archivo = paste(n, "-2", ".top1.csv", sep = "")

Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/", Archivo, sep = "")

write.csv(Top1_Post_FFPE, Archivo_2, row.names = FALSE )

}

```

- **Script para la obtención del espacio clonal ocupado por ese top 1% en tejido:**

```
#Cálculo clonal space del top 1% de los clones en las muestras de FFPE
```

```
#ARCHIVOS PRE
```

```
#Generación de la tabla para los datos a recoger
```

```
Pre_CS_FFPE = data.frame(matrix(nrow = 1, ncol = 1))
```

```
colnames(Pre_CS_FFPE) = c("Patient")
```

```
Pre_CS_FFPE$Extraction = c(NA)
```

```
Pre_CS_FFPE$Tumor = c(NA)
```

```
Pre_CS_FFPE$Response = c(NA)
```

```
Pre_CS_FFPE$Complete_Response = c(NA)
```

```

Pre_CS_FFPE$N_Clones = c(NA)
Pre_CS_FFPE$Clonal_Space = c(NA)

#Lectura de los archivos
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_Pre_1 = "-1.top1.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

  Pre_FFPE=read.csv(Ruta_Pre_2, header = TRUE)

  #Introducción de los datos
  Pre_CS_FFPE[n,1] = n
  Pre_CS_FFPE[n,2] = 1
  Pre_CS_FFPE[n,6] = nrow(Pre_FFPE)
  Pre_CS_FFPE[n,7] = sum(Pre_FFPE$Frequency)

}

#Quito los espacios de los pacientes sin muestra
library("dplyr")
Pre_CS_FFPE = filter(Pre_CS_FFPE, rowSums(is.na(Pre_CS_FFPE)) !=
ncol(Pre_CS_FFPE))

#ARCHIVOS POST
#Generación de la tabla para los datos a recoger
POST_CS_FFPE = data.frame(matrix(nrow = 1, ncol = 1))

```

```

colnames(POST_CS_FFPE) = c("Patient")
POST_CS_FFPE$Extraction = c(NA)
POST_CS_FFPE$Tumor = c(NA)
POST_CS_FFPE$Response = c(NA)
POST_CS_FFPE$Complete_Response = c(NA)
POST_CS_FFPE$N_Clones = c(NA)
POST_CS_FFPE$Clonal_Space = c(NA)

#Lectura de los archivos
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_POST_1 = "-2.top1.csv"
  Ruta_POST_2 = paste(Ruta_General, n, Ruta_POST_1, sep = "")
  if(!file.exists(Ruta_POST_2)) {next}

  POST_FFPE=read.csv(Ruta_POST_2, header = TRUE)

  #Introducción de los datos
  POST_CS_FFPE[n,1] = n
  POST_CS_FFPE[n,2] = 2
  POST_CS_FFPE[n,6] = nrow(POST_FFPE)
  POST_CS_FFPE[n,7] = sum(POST_FFPE$Frequency)

}

#Quito los espacios de los pacientes sin muestra
library("dplyr")
POST_CS_FFPE = filter(POST_CS_FFPE, rowSums(is.na(POST_CS_FFPE)) !=
ncol(POST_CS_FFPE))

#Introduzco datos sobre ganglio o tumor pre

```

```

Ganglio_Pre = c (7, 13, 22, 30, 38, 45, 52)
Pre_CS_FFPE$Tumor = 1
Pre_CS_FFPE[Pre_CS_FFPE$Patient %in% Ganglio_Pre, 3] = 0

#Introduzco datos sobre ganglio o tumor post
Ganglio_Post = c (21, 26, 29, 30, 49)
POST_CS_FFPE$Tumor = 1
POST_CS_FFPE[POST_CS_FFPE$Patient %in% Ganglio_Post, 3] = 0

#Junto los datos de ambos momentos y ordeno los datos
TOP1_CS_FFPE <- rbind(Pre_CS_FFPE, POST_CS_FFPE)
TOP1_CS_FFPE=TOP1_CS_FFPE[order(TOP1_CS_FFPE[,1]),]
rownames(TOP1_CS_FFPE) <- 1:nrow(TOP1_CS_FFPE)

#Introduzco los datos de respuesta
Sin_Respuesta = c(1,18, 31, 44, 52)
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% Sin_Respuesta, 4] = 0
IPR = c(6, 12, 21, 27, 29, 43, 45)
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% IPR, 4] = 1
MPR = c(23, 28, 30, 33, 37, 39, 49, 50)
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% MPR, 4] = 2
CPR = c(2, 4, 7, 9, 11, 13, 14, 15, 16, 17, 19,
        20, 22, 24, 26, 32, 34, 35, 36, 38, 40, 42,
        46, 47, 48, 51)
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% CPR, 4] = 3

TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% CPR, 5] = 1
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% MPR, 5] = 0
TOP1_CS_FFPE[TOP1_CS_FFPE$Patient %in% IPR, 5] = 0

#Genero el archivo con el CS
write.csv(TOP1_CS_FFPE, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Resultados/FFPE/Clonal Space Top 1/CS_Top1_FFPE.csv", row.names = F)

```

- **Script para la obtención de los identificadores de clones y sus frecuencias en sangre:**

```
#Extracción de los clones identificados (variable + nucleotidos) y su frecuencia

#ARCHIVOS PBMCS PRE

#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary/"

  Ruta_Pre_1 = "-1.clone_summary.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

  Pre_PBMCS_CS=read.csv(Ruta_Pre_2, header = TRUE)

  #Creación de la tabla con las filas necesarias
  rows = nrow(Pre_PBMCS_CS)
  Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))
  colnames(Clone_Summary) = c("ID_Clon")
  Clone_Summary$Frequency = c(NA)

  #Introducción de los datos
  Clone_Summary$ID_Clon = paste(Pre_PBMCS_CS$Variable,
Pre_PBMCS_CS$CDR3.NT, sep = " ")
  Clone_Summary$Frequency = Pre_PBMCS_CS$Frequency

  #Guardo la tabla generada
  Archivo = paste(n, "-1", ".clone_summary_analisis.csv", sep = "")
  Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/", Archivo, sep = "")
  write.csv(Clone_Summary, Archivo_2, row.names = FALSE)
}
```


#ARCHIVOS PBMCS POST

#Lectura de los archivos y generación de la tabla con los datos a recoger

for (n in 1:52) {

Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras PBMCS/Clone_Summary/"

Ruta_Post_1 = "-2.clone_summary.csv"

Ruta_Post_2 = paste(Ruta_General, n, Ruta_Post_1, sep = "")

if(!file.exists(Ruta_Post_2)) {next}

Post_PBMCS_CS=read.csv(Ruta_Post_2, header = TRUE)

#Creación de la tabla con las filas necesarias

rows = nrow(Post_PBMCS_CS)

Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))

colnames(Clone_Summary) = c("ID_Clon")

Clone_Summary\$Frequency = c(NA)

#Introducción de los datos

Clone_Summary\$ID_Clon = paste(Post_PBMCS_CS\$Variable, Post_PBMCS_CS\$CDR3.NT, sep = " ")

Clone_Summary\$Frequency = Post_PBMCS_CS\$Frequency

#Guardo la tabla generada

Archivo = paste(n, "-2", ".clone_summary_analisis.csv", sep = "")

Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras PBMCS/Clone_Summary_Analisis/", Archivo, sep = "")

write.csv(Clone_Summary, Archivo_2, row.names = FALSE)

}

#ARCHIVOS PBMCS 6M

#Lectura de los archivos y generación de la tabla con los datos a recoger

for (n in 1:52) {

Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras PBMCS/Clone_Summary/"

```

Ruta_6M_1 = "-3.clone_summary.csv"
Ruta_6M_2 = paste(Ruta_General, n, Ruta_6M_1, sep = "")
if(!file.exists(Ruta_6M_2)) {next}

PBMCS_6M_CS=read.csv(Ruta_6M_2, header = TRUE)

#Creación de la tabla con las filas necesarias
rows = nrow(PBMCS_6M_CS)
Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))
colnames(Clone_Summary) = c("ID_Clon")
Clone_Summary$Frequency = c(NA)

#Introducción de los datos
Clone_Summary$ID_Clon = paste(PBMCS_6M_CS$Variable,
PBMCS_6M_CS$CDR3.NT, sep = " ")
Clone_Summary$Frequency = PBMCS_6M_CS$Frequency

#Guardo la tabla generada
Archivo = paste(n, "-3", ".clone_summary_analisis.csv", sep = "")
Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/", Archivo, sep = "")
write.csv(Clone_Summary, Archivo_2, row.names = FALSE )
}

#ARCHIVOS PBMCS 12M
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {
  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary/"
  Ruta_12M_1 = "-4.clone_summary.csv"
  Ruta_12M_2 = paste(Ruta_General, n, Ruta_12M_1, sep = "")
  if(!file.exists(Ruta_12M_2)) {next}

  PBMCS_12M_CS=read.csv(Ruta_12M_2, header = TRUE)

```

```

#Creación de la tabla con las filas necesarias
rows = nrow(PBMCS_12M_CS)
Clone_Summary = data.frame(matrix(nrow = rows, ncol = 1))
colnames(Clone_Summary) = c("ID_Clon")
Clone_Summary$Frequency = c(NA)

#Introducción de los datos
Clone_Summary$ID_Clon = paste(PBMCS_12M_CS$Variable,
PBMCS_12M_CS$CDR3.NT, sep = " ")
Clone_Summary$Frequency = PBMCS_12M_CS$Frequency

#Guardo la tabla generada
Archivo = paste(n, "-4", ".clone_summary_analisis.csv", sep = "")
Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/", Archivo, sep = "")
write.csv(Clone_Summary, Archivo_2, row.names = FALSE )
}

```

- **Script para la selección de los clones del top 1% en sangre:**

```

#Clones Top 1%

#ARCHIVOS PBMCS PRE
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/"
  Ruta_Pre_1 = "-1.clone_summary_analisis.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

#Calculo el número de clones del top 1 y
Pre_PBMCS_CS=read.csv(Ruta_Pre_2, header = TRUE)

```

```

Pre_PBMCS_CS=Pre_PBMCS_CS[order(Pre_PBMCS_CS[,2], decreasing = TRUE
).]
rownames(Pre_PBMCS_CS) <- 1:nrow(Pre_PBMCS_CS)
Top1 = nrow(Pre_PBMCS_CS) / 100
Top1 = round(Top1)
Top1_Pre_PBMCS = head(Pre_PBMCS_CS, Top1)

#Guardo la tabla generada
Archivo = paste(n, "-1", ".top1.csv", sep = "")
Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/", Archivo, sep = "")
write.csv(Top1_Pre_PBMCS, Archivo_2, row.names = FALSE )

}

```

#ARCHIVOS PBMCS POST

#Lectura de los archivos y generación de la tabla con los datos a recoger

```
for (n in 1:52) {
```

```

Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/"

```

```
Ruta_Post_1 = "-2.clone_summary_analysis.csv"
```

```
Ruta_Post_2 = paste(Ruta_General, n, Ruta_Post_1, sep = "")
```

```
if(!file.exists(Ruta_Post_2)) {next}
```

#Calculo el número de clones del top 1 y

```
Post_PBMCS_CS=read.csv(Ruta_Post_2, header = TRUE)
```

```
Post_PBMCS_CS=Post_PBMCS_CS[order(Post_PBMCS_CS[,2], decreasing =
TRUE ),]
```

```
rownames(Post_PBMCS_CS) <- 1:nrow(Post_PBMCS_CS)
```

```
Top1 = nrow(Post_PBMCS_CS) / 100
```

```
Top1 = round(Top1)
```

```
Top1_Post_PBMCS = head(Post_PBMCS_CS, Top1)
```

```

#Guardo la tabla generada
Archivo = paste(n, "-2", ".top1.csv", sep = "")
Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/", Archivo, sep = "")
write.csv(Top1_Post_PBMCS, Archivo_2, row.names = FALSE )

}

#ARCHIVOS PBMCS 6M
#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/"
  Ruta_6M_1 = "-3.clone_summary_analisis.csv"
  Ruta_6M_2 = paste(Ruta_General, n, Ruta_6M_1, sep = "")
  if(!file.exists(Ruta_6M_2)) {next}

  #Calculo el número de clones del top 1 y
  PBMCS_6M_CS=read.csv(Ruta_6M_2, header = TRUE)
  PBMCS_6M_CS=PBMCS_6M_CS[order(PBMCS_6M_CS[,2], decreasing = TRUE ),]
  rownames(PBMCS_6M_CS) <- 1:nrow(PBMCS_6M_CS)
  Top1 = nrow(PBMCS_6M_CS) / 100
  Top1 = round(Top1)
  Top1_6M_PBMCS = head(PBMCS_6M_CS, Top1)

  #Guardo la tabla generada
  Archivo = paste(n, "-3", ".top1.csv", sep = "")
  Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/", Archivo, sep = "")
  write.csv(Top1_6M_PBMCS, Archivo_2, row.names = FALSE )

}

```

```

#ARCHIVOS PBMCS 12M

#Lectura de los archivos y generación de la tabla con los datos a recoger
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clone_Summary_Analisis/"

  Ruta_12M_1 = "-4.clone_summary_analysis.csv"
  Ruta_12M_2 = paste(Ruta_General, n, Ruta_12M_1, sep = "")
  if(!file.exists(Ruta_12M_2)) {next}

  #Calculo el número de clones del top 1 y
  PBMCS_12M_CS=read.csv(Ruta_12M_2, header = TRUE)
  PBMCS_12M_CS=PBMCS_12M_CS[order(PBMCS_12M_CS[,2], decreasing =
TRUE ),]
  rownames(PBMCS_12M_CS) <- 1:nrow(PBMCS_12M_CS)
  Top1 = nrow(PBMCS_12M_CS) / 100
  Top1 = round(Top1)
  Top1_12M_PBMCS = head(PBMCS_12M_CS, Top1)

  #Guardo la tabla generada
  Archivo = paste(n, "-4", ".top1.csv", sep = "")
  Archivo_2 = paste("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/", Archivo, sep = "")
  write.csv(Top1_12M_PBMCS, Archivo_2, row.names = FALSE )

}

```

- **Script para la obtención del espacio clonal de ese top 1% en sangre:**

```
#Cálculo clonal space del top 1% de los clones en las muestras de PBMCS
```

```
#ARCHIVOS PRE
```

```
#Generación de la tabla para los datos a recoger
```

```

Pre_CS_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(Pre_CS_PBMCS) = c("Patient")
Pre_CS_PBMCS$Extraction = c(NA)
Pre_CS_PBMCS$Response = c(NA)
Pre_CS_PBMCS$Complete_Response = c(NA)
Pre_CS_PBMCS$N_Clones = c(NA)
Pre_CS_PBMCS$Clonal_Space = c(NA)

#Lectura de los archivos
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/"
  Ruta_Pre_1 = "-1.top1.csv"
  Ruta_Pre_2 = paste(Ruta_General, n, Ruta_Pre_1, sep = "")
  if(!file.exists(Ruta_Pre_2)) {next}

  Pre_PBMCS=read.csv(Ruta_Pre_2, header = TRUE)

  #Introducción de los datos
  Pre_CS_PBMCS[n,1] = n
  Pre_CS_PBMCS[n,2] = 1
  Pre_CS_PBMCS[n,5] = nrow(Pre_PBMCS)
  Pre_CS_PBMCS[n,6] = sum(Pre_PBMCS$Frequency)

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")
Pre_CS_PBMCS = filter(Pre_CS_PBMCS, rowSums(is.na(Pre_CS_PBMCS)) !=
ncol(Pre_CS_PBMCS))

```

```

#ARCHIVOS POST
#Generación de la tabla para los datos a recoger
POST_CS_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(POST_CS_PBMCS) = c("Patient")
POST_CS_PBMCS$Extraction = c(NA)
POST_CS_PBMCS$Response = c(NA)
POST_CS_PBMCS$Complete_Response = c(NA)
POST_CS_PBMCS$N_Clones = c(NA)
POST_CS_PBMCS$Clonal_Space = c(NA)

#Lectura de los archivos
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/"
  Ruta_POST_1 = "-2.top1.csv"
  Ruta_POST_2 = paste(Ruta_General, n, Ruta_POST_1, sep = "")
  if(!file.exists(Ruta_POST_2)) {next}

  POST_PBMCS=read.csv(Ruta_POST_2, header = TRUE)

  #Introducción de los datos
  POST_CS_PBMCS[n,1] = n
  POST_CS_PBMCS[n,2] = 2
  POST_CS_PBMCS[n,5] = nrow(POST_PBMCS)
  POST_CS_PBMCS[n,6] = sum(POST_PBMCS$Frequency)

}

#Quito los huecos de los pacientes sin muestra
library("dplyr")

```



```
POST_CS_PBMCS = filter(POST_CS_PBMCS, rowSums(is.na(POST_CS_PBMCS))
!= ncol(POST_CS_PBMCS))
```

```
#ARCHIVOS 6M
```

```
#Generación de la tabla para los datos a recoger
```

```
CS_PBMCS_6M = data.frame(matrix(nrow = 1, ncol = 1))
```

```
colnames(CS_PBMCS_6M) = c("Patient")
```

```
CS_PBMCS_6M$Extraction = c(NA)
```

```
CS_PBMCS_6M$Response = c(NA)
```

```
CS_PBMCS_6M$Complete_Response = c(NA)
```

```
CS_PBMCS_6M$N_Clones = c(NA)
```

```
CS_PBMCS_6M$Clonal_Space = c(NA)
```

```
#Lectura de los archivos
```

```
for (n in 1:52) {
```

```
  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/"
```

```
  Ruta_6M_1 = "-3.top1.csv"
```

```
  Ruta_6M_2 = paste(Ruta_General, n, Ruta_6M_1, sep = "")
```

```
  if(!file.exists(Ruta_6M_2)) {next}
```

```
  PBMCS_6M=read.csv(Ruta_6M_2, header = TRUE)
```

```
#Introducción de los datos
```

```
CS_PBMCS_6M[n,1] = n
```

```
CS_PBMCS_6M[n,2] = 3
```

```
CS_PBMCS_6M[n,5] = nrow(PBMCS_6M)
```

```
CS_PBMCS_6M[n,6] = sum(PBMCS_6M$Frequency)
```

```
}
```

```
#Quito los huecos de los pacientes sin muestra
```

```

library("dplyr")

CS_PBMCS_6M = filter(CS_PBMCS_6M, rowSums(is.na(CS_PBMCS_6M)) !=
ncol(CS_PBMCS_6M))

#ARCHIVOS 12M
#Generación de la tabla para los datos a recoger
CS_PBMCS_12M = data.frame(matrix(nrow = 1, ncol = 1))
colnames(CS_PBMCS_12M) = c("Patient")
CS_PBMCS_12M$Extraction = c(NA)
CS_PBMCS_12M$Response = c(NA)
CS_PBMCS_12M$Complete_Response = c(NA)
CS_PBMCS_12M$N_Clones = c(NA)
CS_PBMCS_12M$Clonal_Space = c(NA)

#Lectura de los archivos
for (n in 1:52) {

  Ruta_General= "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
PBMCS/Clones_Top1/"
  Ruta_12M_1 = "-4.top1.csv"
  Ruta_12M_2 = paste(Ruta_General, n, Ruta_12M_1, sep = "")
  if(!file.exists(Ruta_12M_2)) {next}

  PBMCS_12M=read.csv(Ruta_12M_2, header = TRUE)

  #Introducción de los datos
  CS_PBMCS_12M[n,1] = n
  CS_PBMCS_12M[n,2] = 4
  CS_PBMCS_12M[n,5] = nrow(PBMCS_12M)
  CS_PBMCS_12M[n,6] = sum(PBMCS_12M$Frequency)

}

```

```

#Quito los huecos de los pacientes sin muestra
library("dplyr")
CS_PBMCS_12M = filter(CS_PBMCS_12M, rowSums(is.na(CS_PBMCS_12M)) !=
ncol(CS_PBMCS_12M))

#Junto los datos de ambos momentos y ordeno los datos
TOP1_CS_PBMCS <- rbind(Pre_CS_PBMCS, POST_CS_PBMCS, CS_PBMCS_6M,
CS_PBMCS_12M)
TOP1_CS_PBMCS=TOP1_CS_PBMCS[order(TOP1_CS_PBMCS[,1]),]
rownames(TOP1_CS_PBMCS) <- 1:nrow(TOP1_CS_PBMCS)

#Introduzco los datos de respuesta
Sin_Respuesta = c(1,18, 31, 44, 52)
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% Sin_Respuesta, 3] = 0
IPR = c(6, 12, 21, 27, 29, 43, 45)
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% IPR, 3] = 1
MPR = c(23, 28, 30, 33, 37, 39, 49, 50)
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% MPR, 3] = 2
CPR = c(2, 4, 7, 9, 11, 13, 14, 15, 16, 17, 19,
20, 22, 24, 26, 32, 34, 35, 36, 38, 40, 42,
46, 47, 48, 51)
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% CPR, 3] = 3

TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% CPR, 4] = 1
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% MPR, 4] = 0
TOP1_CS_PBMCS[TOP1_CS_PBMCS$Patient %in% IPR, 4] = 0

#Genero el archivo con el CS
write.csv(TOP1_CS_PBMCS, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Resultados/PBMCS/Clonal Space Top 1/CS_Top1_PBMCS.csv", row.names = F)

```

- **Script para el seguimiento del espacio clonal del top 1% de la muestra basal tisular:**

#Seguimiento de los clones del top 1% de las muestras pre FFPE en el resto de muestras (CLONAL SPACE)

#Comparación con la muestra post de FFPE

#Generación de la tabla con el recopilatorio de los datos

```
CS_Post_FFPE = data.frame(matrix(nrow = 1, ncol = 1))
```

```
colnames(CS_Post_FFPE) = c("Patient")
```

```
CS_Post_FFPE$N_Clones_Seguidos = c(NA)
```

```
CS_Post_FFPE$Freq_Top1PreF_Pre_FFPE = c(NA)
```

```
CS_Post_FFPE$Freq_Top1PreF_Post_FFPE = c(NA)
```

```
for (n in 1:52) {
```

```
  #Lectura del archuivo con los clones top1% de la muestra pre FFPE
```

```
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras FFPE/Clones_Top1/"
```

```
  Ruta_Pre_FFPE_2 = "-1.top1.csv"
```

```
  Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
```

```
  if(!file.exists(Ruta_Pre_FFPE)) {
```

```
    next
```

```
  }
```

```
  Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)
```

```
  #Lectura del archivo con los clones de la muestra post FFPE
```

```
  Ruta_Post_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras FFPE/Clone_Summary_Analisis/"
```

```
  Ruta_Post_FFPE_2 = "-2.clone_summary_analisis.csv"
```

```
  Ruta_Post_FFPE = paste(Ruta_Post_FFPE_1, n, Ruta_Post_FFPE_2, sep = "")
```

```
  if(!file.exists(Ruta_Post_FFPE)) {
```

```
    next
```

```
  }
```

```

Clones_Post_FFPE = read.csv(Ruta_Post_FFPE, header = TRUE)

#Generación de la tabla con los datos a generar de la comparación
Seg_Clones = data.frame(matrix(nrow = nrow(Top1_Pre_FFPE), ncol = 1))
colnames(Seg_Clones) = c("IDs_Pre")
Seg_Clones$Freq_Pre = c(NA)

#Relleno con la frecuencia y los id de pre
Seg_Clones$IDs_Pre = Top1_Pre_FFPE$ID_Clon
Seg_Clones$Freq_Pre = Top1_Pre_FFPE$Frequency

#Relleno con la frecuencia y los id de post (añado las filas suficientes y dejo las
sobrantes de pre como NA)
Seg_Clones$IDs_Post = c(NA)
Seg_Clones$Freq_Post = c(NA)

rowsPost = nrow(Clones_Post_FFPE)
Seg_Clones[rowsPost,] <- NA

Seg_Clones$IDs_Post = Clones_Post_FFPE$ID_Clon
Seg_Clones$Freq_Post = Clones_Post_FFPE$Frequency

#Búsqueda de los clones de pre en las muestras de post
Seg_Clones$ClonPre_en_Post = ifelse(Seg_Clones$IDs_Pre %in%
Seg_Clones$IDs_Post, 1, 0)

#Generación de la variable de la freq del clon pre en muestra post
Seg_Clones$Freq_ClonPre_MuestraPost = c(0)
for (x in 1:nrow(Seg_Clones)) {
  y = which(Seg_Clones$IDs_Post == Seg_Clones$IDs_Pre[x])
  if (length(y)>0){
    Seg_Clones$Freq_ClonPre_MuestraPost[x] = Seg_Clones$Freq_Post[y]
  }
}

```

```

#Introduzco los datos obtenidos de la comparación
CS_Post_FFPE[n,1] = n
CS_Post_FFPE[n,2] = nrow(Top1_Pre_FFPE)
CS_Post_FFPE[n,3] = sum(Top1_Pre_FFPE$Frequency)
CS_Post_FFPE[n,4] = sum(Seg_Clones$Freq_ClonPre_MuestraPost)

}

CS_Post_FFPE[52,] = NA

#
#

#Comparación con la muestra pre de PBMCS
#Generación de la tabla con el recopilatorio de los datos
CS_Pre_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(CS_Pre_PBMCS) = c("Patient")
CS_Pre_PBMCS$N_Clones_Seguidos = c(NA)
CS_Pre_PBMCS$Freq_Top1PreF_Pre_FFPE = c(NA)
CS_Pre_PBMCS$Freq_Top1PreF_Pre_PBMCS = c(NA)

for (n in 1:52) {
  #Lectura del archiivo con los clones top1% de la muestra pre FFPE
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_Pre_FFPE_2 = "-1.top1.csv"
  Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
  if(!file.exists(Ruta_Pre_FFPE)) {
    next
  }
  Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)

```

```

#Lectura del archivo con los clones de la muestra pre PBMCS
Ruta_Pre_PBMCS_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Muestras PBMCS/Clone_Summary_Analisis/"
Ruta_Pre_PBMCS_2 = "-1.clone_summary_analisis.csv"
Ruta_Pre_PBMCS = paste(Ruta_Pre_PBMCS_1, n, Ruta_Pre_PBMCS_2, sep = "")
if(!file.exists(Ruta_Pre_PBMCS)) {
  next
}
Clones_Pre_PBMCS = read.csv(Ruta_Pre_PBMCS, header = TRUE)

#Generación de la tabla con los datos a generar de la comparación
Seg_Clones = data.frame(matrix(nrow = nrow(Top1_Pre_FFPE), ncol = 1))
colnames(Seg_Clones) = c("IDs_Pre_FFPE")
Seg_Clones$Freq_Pre_FFPE = c(NA)

#Relleno con la frecuencia y los id de pre FFPE
Seg_Clones$IDs_Pre_FFPE = Top1_Pre_FFPE$ID_Clon
Seg_Clones$Freq_Pre_FFPE = Top1_Pre_FFPE$Frequency

#Relleno con la frecuencia y los id de pre PBMCS (añado las filas suficientes y dejo
las sobrantes de pre como NA)
Seg_Clones$IDs_Pre_PBMCS = c(NA)
Seg_Clones$Freq_Pre_PBMCS = c(NA)

rowsPrePBMCS = nrow(Clones_Pre_PBMCS)
Seg_Clones[rowsPrePBMCS,] <- NA

Seg_Clones$IDs_Pre_PBMCS = Clones_Pre_PBMCS$ID_Clon
Seg_Clones$Freq_Pre_PBMCS = Clones_Pre_PBMCS$Frequency

#Búsqueda de los clones de pre FFPE en las muestras de pre PBMCS
Seg_Clones$ClonPreF_en_PreP = ifelse(Seg_Clones$IDs_Pre_FFPE %in%
Seg_Clones$IDs_Pre_PBMCS, 1, 0)

```

```

#Generación de la variable de la freq del clon pre FFPE en muestra pre PBMCS
Seg_Clones$Freq_ClonPreF_MuestraPreP = c(0)
for (x in 1:nrow(Seg_Clones)) {
  y = which(Seg_Clones$IDs_Pre_PBMCS == Seg_Clones$IDs_Pre_FFPE[x])
  if (length(y)>0){
    Seg_Clones$Freq_ClonPreF_MuestraPreP[x] = Seg_Clones$Freq_Pre_PBMCS[y]
  }
}

#Introduzco los datos obtenidos de la comparación
CS_Pre_PBMCS[n,1] = n
CS_Pre_PBMCS[n,2] = nrow(Top1_Pre_FFPE)
CS_Pre_PBMCS[n,3] = sum(Top1_Pre_FFPE$Frequency)
CS_Pre_PBMCS[n,4] = sum(Seg_Clones$Freq_ClonPreF_MuestraPreP)

}

#
#

#Comparación con la muestra post de PBMCS
#Generación de la tabla con el recopilatorio de los datos
CS_Post_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(CS_Post_PBMCS) = c("Patient")
CS_Post_PBMCS$N_Clones_Seguidos = c(NA)
CS_Post_PBMCS$Freq_Top1PreF_Pre_FFPE = c(NA)
CS_Post_PBMCS$Freq_Top1PreF_Post_PBMCS = c(NA)

for (n in 1:52) {
  #Lectura del archivo con los clones top1% de la muestra pre FFPE
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"

```



```

Ruta_Pre_FFPE_2 = "-1.top1.csv"
Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
if(!file.exists(Ruta_Pre_FFPE)) {
  next
}
Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)

#Lectura del archivo con los clones de la muestra post PBMCS
Ruta_Post_PBMCS_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Muestras PBMCS/Clone_Summary_Analisis/"
Ruta_Post_PBMCS_2 = "-2.clone_summary_analisis.csv"
Ruta_Post_PBMCS = paste(Ruta_Post_PBMCS_1, n, Ruta_Post_PBMCS_2, sep =
"")
if(!file.exists(Ruta_Post_PBMCS)) {
  next
}
Clones_Post_PBMCS = read.csv(Ruta_Post_PBMCS, header = TRUE)

#Generación de la tabla con los datos a generar de la comparación
Seg_Clones = data.frame(matrix(nrow = nrow(Top1_Pre_FFPE), ncol = 1))
colnames(Seg_Clones) = c("IDs_Pre_FFPE")
Seg_Clones$Freq_Pre_FFPE = c(NA)

#Relleno con la frecuencia y los id de pre FFPE
Seg_Clones$IDs_Pre_FFPE = Top1_Pre_FFPE$ID_Clon
Seg_Clones$Freq_Pre_FFPE = Top1_Pre_FFPE$Frequency

#Relleno con la frecuencia y los id de post PBMCS (añado las filas suficientes y dejo
las sobrantes de pre como NA)
Seg_Clones$IDs_Post_PBMCS = c(NA)
Seg_Clones$Freq_Post_PBMCS = c(NA)

rowsPostPBMCS = nrow(Clones_Post_PBMCS)
Seg_Clones[rowsPostPBMCS,] <- NA

```

```

Seg_Clones$IDs_Post_PBMCS = Clones_Post_PBMCS$ID_Clon
Seg_Clones$Freq_Post_PBMCS = Clones_Post_PBMCS$Frequency

#Búsqueda de los clones de pre en las muestras de post
Seg_Clones$ClonPreF_en_PostP = ifelse(Seg_Clones$IDs_Pre_FFPE %in%
Seg_Clones$IDs_Post_PBMCS, 1, 0)

#Generación de la variable de la freq del clon pre en muestra post
Seg_Clones$Freq_ClonPreF_MuestraPostP = c(0)
for (x in 1:nrow(Seg_Clones)) {
  y = which(Seg_Clones$IDs_Post_PBMCS == Seg_Clones$IDs_Pre_FFPE[x])
  if (length(y)>0){
    Seg_Clones$Freq_ClonPreF_MuestraPostP[x] =
Seg_Clones$Freq_Post_PBMCS[y]
  }
}

#Introduzco los datos obtenidos de la comparación
CS_Post_PBMCS[n,1] = n
CS_Post_PBMCS[n,2] = nrow(Top1_Pre_FFPE)
CS_Post_PBMCS[n,3] = sum(Top1_Pre_FFPE$Frequency)
CS_Post_PBMCS[n,4] = sum(Seg_Clones$Freq_ClonPreF_MuestraPostP)

}

#
#

#Comparación con la muestra 6M de PBMCS
#Generación de la tabla con el recopilatorio de los datos
CS_6M_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(CS_6M_PBMCS) = c("Patient")
CS_6M_PBMCS$N_Clones_Seguidos = c(NA)

```

```

CS_6M_PBMCS$Freq_Top1PreF_Pre_FFPE = c(NA)
CS_6M_PBMCS$Freq_Top1PreF_6M_PBMCS = c(NA)

for (n in 1:52) {
  #Lectura del archuivo con los clones top1% de la muestra pre FFPE
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_Pre_FFPE_2 = "-1.top1.csv"
  Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
  if(!file.exists(Ruta_Pre_FFPE)) {
    next
  }
  Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)

  #Lectura del archivo con los clones de la muestra 6M PBMCS
  Ruta_6M_PBMCS_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Muestras PBMCS/Clone_Summary_Analisis/"
  Ruta_6M_PBMCS_2 = "-3.clone_summary_analisis.csv"
  Ruta_6M_PBMCS = paste(Ruta_6M_PBMCS_1, n, Ruta_6M_PBMCS_2, sep = "")
  if(!file.exists(Ruta_6M_PBMCS)) {
    next
  }
  Clones_6M_PBMCS = read.csv(Ruta_6M_PBMCS, header = TRUE)

  #Generación de la tabla con los datos a generar de la comparación
  Seg_Clones = data.frame(matrix(nrow = nrow(Top1_Pre_FFPE), ncol = 1))
  colnames(Seg_Clones) = c("IDs_Pre_FFPE")
  Seg_Clones$Freq_Pre_FFPE = c(NA)

  #Relleno con la frecuencia y los id de pre FFPE
  Seg_Clones$IDs_Pre_FFPE = Top1_Pre_FFPE$ID_Clon
  Seg_Clones$Freq_Pre_FFPE = Top1_Pre_FFPE$Frequency

```

```
#Relleno con la frecuencia y los id de post PBMCS (añado las filas suficientes y dejo las sobrantes de pre como NA)
```

```
Seg_Clones$IDs_6M_PBMCS = c(NA)
```

```
Seg_Clones$Freq_6M_PBMCS = c(NA)
```

```
rows6MPBMCS = nrow(Clones_6M_PBMCS)
```

```
Seg_Clones[rows6MPBMCS,] <- NA
```

```
Seg_Clones$IDs_6M_PBMCS = Clones_6M_PBMCS$ID_Clon
```

```
Seg_Clones$Freq_6M_PBMCS = Clones_6M_PBMCS$Frequency
```

```
#Búsqueda de los clones de pre en las muestras de 6M
```

```
Seg_Clones$ClonPreF_en_6MP = ifelse(Seg_Clones$IDs_Pre_FFPE %in%  
Seg_Clones$IDs_6M_PBMCS, 1, 0)
```

```
#Generación de la variable de la freq del clon pre en muestra 6M
```

```
Seg_Clones$Freq_ClonPreF_Muestra6MP = c(0)
```

```
for (x in 1:nrow(Seg_Clones)) {
```

```
  y = which(Seg_Clones$IDs_6M_PBMCS == Seg_Clones$IDs_Pre_FFPE[x])
```

```
  if (length(y)>0){
```

```
    Seg_Clones$Freq_ClonPreF_Muestra6MP[x] = Seg_Clones$Freq_6M_PBMCS[y]
```

```
  }}
```

```
#Introduzco los datos obtenidos de la comparación
```

```
CS_6M_PBMCS[n,1] = n
```

```
CS_6M_PBMCS[n,2] = nrow(Top1_Pre_FFPE)
```

```
CS_6M_PBMCS[n,3] = sum(Top1_Pre_FFPE$Frequency)
```

```
CS_6M_PBMCS[n,4] = sum(Seg_Clones$Freq_ClonPreF_Muestra6MP)
```

```
}
```

```
CS_6M_PBMCS[52,] = NA
```

```

#
#

#Comparación con la muestra 12M de PBMCS
#Generación de la tabla con el recopilatorio de los datos
CS_12M_PBMCS = data.frame(matrix(nrow = 1, ncol = 1))
colnames(CS_12M_PBMCS) = c("Patient")
CS_12M_PBMCS$N_Clones_Seguidos = c(NA)
CS_12M_PBMCS$Freq_Top1PreF_Pre_FFPE = c(NA)
CS_12M_PBMCS$Freq_Top1PreF_12M_PBMCS = c(NA)

for (n in 1:52) {
  #Lectura del archuivo con los clones top1% de la muestra pre FFPE
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_Pre_FFPE_2 = "-1.top1.csv"
  Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
  if(!file.exists(Ruta_Pre_FFPE)) {
    next
  }
  Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)

  #Lectura del archivo con los clones de la muestra 12M PBMCS
  Ruta_12M_PBMCS_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Muestras PBMCS/Clone_Summary_Analisis/"
  Ruta_12M_PBMCS_2 = "-4.clone_summary_analisis.csv"
  Ruta_12M_PBMCS = paste(Ruta_12M_PBMCS_1, n, Ruta_12M_PBMCS_2, sep =
"")
  if(!file.exists(Ruta_12M_PBMCS)) {
    next
  }
  Clones_12M_PBMCS = read.csv(Ruta_12M_PBMCS, header = TRUE)
}

```

```

#Generación de la tabla con los datos a generar de la comparación
Seg_Clones = data.frame(matrix(nrow = nrow(Top1_Pre_FFPE), ncol = 1))
colnames(Seg_Clones) = c("IDs_Pre_FFPE")
Seg_Clones$Freq_Pre_FFPE = c(NA)

#Relleno con la frecuencia y los id de pre FFPE
Seg_Clones$IDs_Pre_FFPE = Top1_Pre_FFPE$ID_Clon
Seg_Clones$Freq_Pre_FFPE = Top1_Pre_FFPE$Frequency

#Relleno con la frecuencia y los id de 12M PBMCS (añado las filas suficientes y dejo
las sobrantes de pre como NA)
Seg_Clones$IDs_12M_PBMCS = c(NA)
Seg_Clones$Freq_12M_PBMCS = c(NA)

rows12MPBMCS = nrow(Clones_12M_PBMCS)
Seg_Clones[rows12MPBMCS,] <- NA

Seg_Clones$IDs_12M_PBMCS = Clones_12M_PBMCS$ID_Clon
Seg_Clones$Freq_12M_PBMCS = Clones_12M_PBMCS$Frequency

#Búsqueda de los clones de pre en las muestras de 12M
Seg_Clones$ClonPreF_en_12MP = ifelse(Seg_Clones$IDs_Pre_FFPE %in%
Seg_Clones$IDs_12M_PBMCS, 1, 0)

#Generación de la variable de la freq del clon pre en muestra 12M
Seg_Clones$Freq_ClonPreF_Muestra12MP = c(0)
for (x in 1:nrow(Seg_Clones)) {
  y = which(Seg_Clones$IDs_12M_PBMCS == Seg_Clones$IDs_Pre_FFPE[x])
  if (length(y)>0){
    Seg_Clones$Freq_ClonPreF_Muestra12MP[x] =
Seg_Clones$Freq_12M_PBMCS[y]
  }
}

```

```

#Introduzco los datos obtenidos de la comparación
CS_12M_PBMCS[n,1] = n
CS_12M_PBMCS[n,2] = nrow(Top1_Pre_FFPE)
CS_12M_PBMCS[n,3] = sum(Top1_Pre_FFPE$Frequency)
CS_12M_PBMCS[n,4] = sum(Seg_Clones$Freq_ClonPreF_Muestra12MP)

}

CS_12M_PBMCS[52,] = NA

#
#

#Combinación de las tablas recopilatorias
Seg_CS_Top1_Pre_FFPE = data.frame(matrix(nrow = 1, ncol = 1))
colnames(Seg_CS_Top1_Pre_FFPE) = c("Patient")
Seg_CS_Top1_Pre_FFPE$Response = c(NA)
Seg_CS_Top1_Pre_FFPE$Complete_Response = c(NA)
Seg_CS_Top1_Pre_FFPE$N_Clones_Seguidos = c(NA)
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_Pre_FFPE = c(NA)
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_Post_FFPE = c(NA)
for (n in 1:52) {
  #Lectura del archuivo con los clones top1% de la muestra pre FFPE
  Ruta_Pre_FFPE_1 = "C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Muestras
FFPE/Clones_Top1/"
  Ruta_Pre_FFPE_2 = "-1.top1.csv"
  Ruta_Pre_FFPE = paste(Ruta_Pre_FFPE_1, n, Ruta_Pre_FFPE_2, sep = "")
  if(!file.exists(Ruta_Pre_FFPE)) {
    next
  }
  Top1_Pre_FFPE = read.csv(Ruta_Pre_FFPE, header = TRUE)

#Introducción de los datos

```

```

Seg_CS_Top1_Pre_FFPE[n,1] = n
Seg_CS_Top1_Pre_FFPE[n,5] = sum(Top1_Pre_FFPE$Frequency)
Seg_CS_Top1_Pre_FFPE[n,4] = nrow(Top1_Pre_FFPE)

}

Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_Post_FFPE =
CS_Post_FFPE$Freq_Top1PreF_Post_FFPE
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_Pre_PBMCS =
CS_Pre_PBMCS$Freq_Top1PreF_Pre_PBMCS
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_Post_PBMCS =
CS_Post_PBMCS$Freq_Top1PreF_Post_PBMCS
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_6M_PBMCS =
CS_6M_PBMCS$Freq_Top1PreF_6M_PBMCS
Seg_CS_Top1_Pre_FFPE$CS_ClonesTop1_12M_PBMCS =
CS_12M_PBMCS$Freq_Top1PreF_12M_PBMCS
Seg_CS_Top1_Pre_FFPE$Patient = c(1:52)

#Meto los datos de respuesta
Pacientes_Respuesta_1 = c(6, 12, 21, 27, 29, 43, 45)
Seg_CS_Top1_Pre_FFPE[Pacientes_Respuesta_1,2] = 1
Pacientes_Respuesta_2 = c(23, 28, 30, 33, 37, 39, 49, 50)
Seg_CS_Top1_Pre_FFPE[Pacientes_Respuesta_2,2] = 2
Pacientes_Respuesta_3 = c(2, 4, 7, 9, 11, 13, 14, 15, 16, 17, 19, 20, 22, 24, 26, 32,
34, 35, 36, 38, 40, 42, 46, 47, 48, 51)
Seg_CS_Top1_Pre_FFPE[Pacientes_Respuesta_3,2] = 3
Seg_CS_Top1_Pre_FFPE$Complete_Response =
ifelse(Seg_CS_Top1_Pre_FFPE$Response == 3, 1, 0)
library("tidyverse")
Seg_CS_Top1_Pre_FFPE = Seg_CS_Top1_Pre_FFPE %>%
drop_na(N_Clones_Seguidos)

#Guardo los resultados
write.csv(Seg_CS_Top1_Pre_FFPE, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Resultados/FFPE/Seguimiento_Clones_Top1_Pre/Seguimiento_CSTop1_Pre_FF
PE.csv", row.names = F)

```


- **Script para el seguimiento del espacio clonal del top 1% de la muestra basal post-neoadyuvancia:**

#Análisis estadístico del seguimiento del CS del top1% de la muestra post FFPE en el resto de muestras

#Leo el archivo con los datos a analizar

```
Seg_CSTop1_Post_FFPE = read.csv("C:/Users/cmart/Desktop/PDH/TFM Seguimiento TCR/Resultados/FFPE/Seguimiento_Clones_Top1_Post/Seguimiento_CSTop1_Post_FFPE.csv", header = TRUE)
```

#Análisis con respuesta CPR vs NCPR

#Filtro y ordeno los datos

```
library("tidyverse")
```

```
Respuesta = Seg_CSTop1_Post_FFPE %>% drop_na(Complete_Response)
```

```
Respuesta = Respuesta[order(Respuesta[,3]),]
```

```
rownames(Respuesta) <- 1:nrow(Respuesta)
```

#Genero una tabla para recopilar los datos del análisis

```
col = colnames(Respuesta)[6:length(colnames(Respuesta))]
```

```
RE_1 = as.data.frame(col)
```

```
colnames(RE_1) = c("Variable")
```

```
RE_1$P_VALUE = c(NA)
```

```
RE_1$P_VALUE_ADJ = c(NA)
```

```
RE_1$Median_CPR = c(NA)
```

```
RE_1$Median_NCPR = c(NA)
```

```
RE_1$N_CPR = c(NA)
```

```
RE_1$N_NCPR = c(NA)
```

```
RE_1$P25_CPR = c(NA)
```

```
RE_1$P75_CPR = c(NA)
```

```
RE_1$P25_NCPR = c(NA)
```

```
RE_1$P75_NCPR = c(NA)
```

#Ejecuto el análisis

```
for (l in 6:ncol(Respuesta)) {
```

```

x=wilcox.test(na.omit(Respuesta[Respuesta$Complete_Response == 0, I]),
na.omit(Respuesta[Respuesta$Complete_Response == 1, I]))

pv = x$p.value

RE_1$P_VALUE[I-5] = pv

RE_1$N_NCPR[I-5] = length(na.omit(Respuesta[Respuesta$Complete_Response ==
0, I]))

RE_1$N_CPR[I-5] = length(na.omit(Respuesta[Respuesta$Complete_Response ==
1, I]))

RE_1$Median_NCPR[I-5] =
median(na.omit(Respuesta[Respuesta$Complete_Response == 0,I]))

RE_1$Median_CPR[I-5] =
median(na.omit(Respuesta[Respuesta$Complete_Response == 1,I]))

RE_1$P25_NCPR[I-5] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 0, I]), 0.25)

RE_1$P75_NCPR[I-5] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 0, I]), 0.75)

RE_1$P25_CPR[I-5] = quantile(na.omit(Respuesta[Respuesta$Complete_Response
== 1, I]), 0.25)

RE_1$P75_CPR[I-5] = quantile(na.omit(Respuesta[Respuesta$Complete_Response
== 1, I]), 0.75)

}

```

```

padj = p.adjust(RE_1$P_VALUE, method = "bonferroni", n = length(RE_1$Variable))
RE_1$P_VALUE_ADJ = padj

```

#Guardo los resultados

```

write.csv(RE_1, "C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Resultados/FFPE/Seguimiento_Clones_Top1_Post/Analisis_Seguimiento_CSTop
1_Post_FFPE.csv", row.names = F)

```

- **Script modelo para el análisis U de Mann Whitney:**

```
#Análisis estadístico U Mann Whitney
```

```
#Leo el archivo con los datos a analizar
```

```
Datos = read.csv("Ruta del archivo", header = TRUE)
```

```
#Análisis con respuesta CPR vs NCPR
```

```
#Filtro y ordeno los datos
```

```

library("tidyverse")

Respuesta = Datos %>% drop_na(Complete_Response)

Respuesta = Respuesta[order(Respuesta[,columna con datos de respuesta],)
rownames(Respuesta) <- 1:nrow(Respuesta)

#Genero una tabla para recopilar los datos del análisis

col = colnames(Respuesta)[columna de inicio con datos a analizar
:length(colnames(Respuesta))]

RE_1 = as.data.frame(col)

colnames(RE_1) = c("Variable")

RE_1$P_VALUE = c(NA)
RE_1$P_VALUE_ADJ = c(NA)
RE_1$Median_CPR = c(NA)
RE_1$Median_NCPR = c(NA)
RE_1$N_CPR = c(NA)
RE_1$N_NCPR = c(NA)
RE_1$P25_CPR = c(NA)
RE_1$P75_CPR = c(NA)
RE_1$P25_NCPR = c(NA)
RE_1$P75_NCPR = c(NA)

#Ejecuto el análisis

for (l in columna de inicio con datos a analizar:ncol(Respuesta)) {
  x=wilcox.test(na.omit(Respuesta[Respuesta$Complete_Response == 0, l]),
na.omit(Respuesta[Respuesta$Complete_Response == 1, l]))

  pv = x$p.value

  RE_1$P_VALUE[l-(1- columna de inicio con datos a analizar)] = pv

  RE_1$N_NCPR[l-(1- columna de inicio con datos a analizar)] =
length(na.omit(Respuesta[Respuesta$Complete_Response == 0, l]))

  RE_1$N_CPR[l-(1- columna de inicio con datos a analizar)] =
length(na.omit(Respuesta[Respuesta$Complete_Response == 1, l]))

  RE_1$Median_NCPR[l-(1- columna de inicio con datos a analizar)] =
median(na.omit(Respuesta[Respuesta$Complete_Response == 0,l]))

  RE_1$Median_CPR[l-(1- columna de inicio con datos a analizar)] =
median(na.omit(Respuesta[Respuesta$Complete_Response == 1,l]))

  RE_1$P25_NCPR[l-(1- columna de inicio con datos a analizar)] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 0, l]), 0.25)

```

```

RE_1$P75_NCPRI[-(1- columna de inicio con datos a analizar)] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 0, I]), 0.75)

RE_1$P25_CPRI[-(1- columna de inicio con datos a analizar)] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 1, I]), 0.25)

RE_1$P75_CPRI[-(1- columna de inicio con datos a analizar)] =
quantile(na.omit(Respuesta[Respuesta$Complete_Response == 1, I]), 0.75)

}

padj = p.adjust(RE_1$P_VALUE, method = "bonferroni", n = length(RE_1$Variable))
RE_1$P_VALUE_ADJ = padj

#Guardo los resultados
write.csv(RE_1, "Ruta de destino", row.names = F)

```

- **Script modelo para el análisis de Wilcoxon:**

```

#Quito las muestras no pareadas y ordeno los datos
library(tidyverse)

ALL = subset(Metricas_FFPE_Analisis,duplicated(Metricas_FFPE_Analisis$Patient) |
duplicated(Metricas_FFPE_Analisis$Patient, fromLast=TRUE))

ALL=ALL[order(ALL[,2] ),]

rownames(ALL) <- 1:nrow(ALL)

#Genero una tabla para recopilar los datos del análisis
col = colnames(ALL)[6:length(colnames(ALL))]
RE_5 = as.data.frame(col)
colnames(RE_5) = c("Variable")
RE_5[1,1] = "Clones_ALL"
RE_5[2,1] = "Evenness_ALL"
RE_5[3,1] = "Diversity_ALL"
RE_5[4,1] = "Convergence_ALL"
RE_5$P_VALUE = c(NA)
RE_5$P_VALUE_ADJ = c(NA)
RE_5$Median_PRE = c(NA)
RE_5$Median_POST = c(NA)
RE_5$N_PRE = c(NA)
RE_5$N_POST = c(NA)

```

```
RE_5$P25_PRE = c(NA)
RE_5$P75_PRE = c(NA)
RE_5$P25_POST = c(NA)
RE_5$P75_POST = c(NA)
```

```
#Ejecuto el análisis (Wilcoxon pareado)
```

```
for (l in 6:ncol(ALL)) {
  x=wilcox.test(ALL[ALL$Extraction == 1, l], ALL[ALL$Extraction == 2, l], paired =
TRUE, alternative = "two.sided")
  pv = x$p.value
  RE_5$P_VALUE[l-5] = pv
  RE_5$N_PRE[l-5] = length(na.omit(ALL[ALL$Extraction == 1, l]))
  RE_5$N_POST[l-5] = length(na.omit(ALL[ALL$Extraction == 2, l]))
  RE_5$Median_PRE[l-5] = median(na.omit(ALL[ALL$Extraction == 1,l]))
  RE_5$Median_POST[l-5] = median(na.omit(ALL[ALL$Extraction == 2,l]))
  RE_5$P25_PRE[l-5] = quantile(ALL[ALL$Extraction == 1, l], 0.25)
  RE_5$P75_PRE[l-5] = quantile(ALL[ALL$Extraction == 1, l], 0.75)
  RE_5$P25_POST[l-5] = quantile(ALL[ALL$Extraction == 2, l], 0.25)
  RE_5$P75_POST[l-5] = quantile(ALL[ALL$Extraction == 2, l], 0.75)
}
```

```
RE_5$Median_PRE = format(RE_5$Median_PRE, scientific = FALSE)
RE_5$Median_POST = format(RE_5$Median_POST, scientific = FALSE)
RE_5$P25_PRE = format(RE_5$P25_PRE, scientific = FALSE)
RE_5$P25_POST = format(RE_5$P25_POST, scientific = FALSE)
RE_5$P75_PRE = format(RE_5$P75_PRE, scientific = FALSE)
RE_5$P75_POST = format(RE_5$P75_POST, scientific = FALSE)
```

```
padj = p.adjust(RE_5$P_VALUE, method = "bonferroni", n = length(RE_5$P_VALUE))
RE_5$P_VALUE_ADJ = padj
```

- **Script modelo para el cálculo del Log2 Fold Change:**

```
#Abro el archivo con los datos
```

```
CSTOP1 = read.delim ("C:/Users/cmart/Desktop/PDH/TFM Seguimiento
TCR/Resultados/FFPE/Clonal Space Top 1/CS_Top1_FFPE.csv", header = TRUE,
sep = ",")
```

```
#ANALISIS ALL PACIENTES (PRE VS POST)
```

```
#Quito las muestras no pareadas y ordeno los datos
```

```
library(dplyr)
```

```
ALL_PrevsPost = filter(CSTOP1, Extraction != 3)
```

```
ALL_PrevsPost = filter(ALL_PrevsPost, Extraction != 4)
```

```
ALL_PrevsPost = na.omit(ALL_PrevsPost)
```

```
library(tidyverse)
```

```
ALL_PrevsPost = subset(ALL_PrevsPost,duplicated(ALL_PrevsPost$Patient) |
duplicated(ALL_PrevsPost$Patient, fromLast=TRUE))
```

```
ALL_PrevsPost=ALL_PrevsPost[order(ALL_PrevsPost[,2] ),]
```

```
rownames(ALL_PrevsPost) <- 1:nrow(ALL_PrevsPost)
```

```
#Genero una tabla para recopilar los datos del cálculo
```

```
col = colnames(ALL_PrevsPost)[7:length(colnames(ALL_PrevsPost))]
```

```
FC_1 = as.data.frame(col)
```

```
colnames(FC_1) = c("Variable")
```

```
FC_1[1,1] = "CSTOP1FC"
```

```
FC_1$FOLD_CHANGE = c(NA)
```

```
FC_1$Median_PRE = c(NA)
```

```
FC_1$Median_POST = c(NA)
```

```
FC_1$N_PRE = c(NA)
```

```
FC_1$N_POST = c(NA)
```

```
FC_1$P25_PRE = c(NA)
```

```
FC_1$P75_PRE = c(NA)
```

```
FC_1$P25_POST = c(NA)
```

```
FC_1$P75_POST = c(NA)
```

```
for (l in 7:ncol(ALL_PrevsPost)) {
```

```

x=log2(mean(ALL_PrevsPost[ALL_PrevsPost$Extraction == 2,
l])/mean(ALL_PrevsPost[ALL_PrevsPost$Extraction == 1, l]))
FC_1$FOLD_CHANGE[l-6] = x

FC_1$N_PRE[l-6] = length(na.omit(ALL_PrevsPost[ALL_PrevsPost$Extraction == 1,
l]))
FC_1$N_POST[l-6] = length(na.omit(ALL_PrevsPost[ALL_PrevsPost$Extraction ==
2, l]))
FC_1$Median_PRE[l-6] = median(na.omit(ALL_PrevsPost[ALL_PrevsPost$Extraction
== 1,l]))
FC_1$Median_POST[l-6] =
median(na.omit(ALL_PrevsPost[ALL_PrevsPost$Extraction == 2,l]))
FC_1$P25_PRE[l-6] = quantile(ALL_PrevsPost[ALL_PrevsPost$Extraction == 1, l],
0.25)
FC_1$P75_PRE[l-6] = quantile(ALL_PrevsPost[ALL_PrevsPost$Extraction == 1, l],
0.75)
FC_1$P25_POST[l-6] = quantile(ALL_PrevsPost[ALL_PrevsPost$Extraction == 2, l],
0.25)
FC_1$P75_POST[l-6] = quantile(ALL_PrevsPost[ALL_PrevsPost$Extraction == 2, l],
0.75)

}

```

```

FC_1$Median_PRE = format(FC_1$Median_PRE, scientific = FALSE)
FC_1$Median_POST = format(FC_1$Median_POST, scientific = FALSE)
FC_1$P25_PRE = format(FC_1$P25_PRE, scientific = FALSE)
FC_1$P25_POST = format(FC_1$P25_POST, scientific = FALSE)
FC_1$P75_PRE = format(FC_1$P75_PRE, scientific = FALSE)
FC_1$P75_POST = format(FC_1$P75_POST, scientific = FALSE)

```

Anexo 2. Aprobación del comité ético para el proyecto:

**D^a. CRISTINA AVENDAÑO SOLÁ, PRESIDENTA DEL COMITÉ ETICO DE
INVESTIGACIÓN CON MEDICAMENTOS DEL HOSPITAL
UNIVERSITARIO PUERTA DE HIERRO MAJADAHONDA DE MADRID**

CERTIFICA

Que dicho Comité ha evaluado el proyecto de investigación titulado:

**“MICROAMBIENTE TUMORAL E IDENTIFICACIÓN DE BIOMARCADORES DE
RESPUESTA A QUIMIO-INMUNO TERAPIA NEOADYUVANTE EN PACIENTES
CON CÁNCER DE PULMÓN CÉLULA NO PEQUEÑA”.**

a solicitud de D. Alberto Cruz Bermúdez del Servicio de Oncología Médica y
cuyo Director del proyecto es el Dr. Mariano Provencio del Hospital
Universitario Puerta de Hierro Majadahonda, considerando que su planteamiento
global es aceptable desde el punto de vista metodológico y ético. Acta nº 17.18.

En Majadahonda, a 29 de octubre de 2018


Hospital Universitario
Puerta de Hierro
Majadahonda
SaludMadrid
Comunidad de Madrid
Comité Ético de Investigación
Clínica

Fdo.: Dra. Avendaño Solá
Presidenta del CEIm

DICTAMEN DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS

D^a. **Cristina Avendaño Solá**, Presidente del Comité de Ética de la Investigación con medicamentos DEL HOSPITAL UNIVERSITARIO PUERTA DE HIERRO MAJADAHONDA

CERTIFICA

Que este Comité ha evaluado la siguiente propuesta del ensayo clínico:

Código de protocolo: **NADIM (GECF 16/03)**

Nº de EudraCT: **2016-003732-20**

Título: **Ensayo clínico fase II, exploratorio y multicéntrico, de quimioinmunoterapia neo-adyuvante para el tratamiento del cáncer de pulmón no microcítico estadio III operable.**

Promotor: **Grupo Español de Cáncer de Pulmón (SLCG/GECF)**

Protocolo: **v 3.6 de 22/12/2016**

HIP/CI Principal: **v.1.2 de 27 Diciembre 2016**

Que este Comité ha realizado la evaluación de la parte I de la solicitud de autorización del ensayo, ha valorado las respuestas del promotor a las aclaraciones solicitadas (si las hubiera) y ha transmitido a la Agencia Española de medicamentos su opinión final sobre la parte I.

Que este Comité ha realizado la evaluación de la parte II de la solicitud de autorización del ensayo, de acuerdo con lo previsto en el Real Decreto 1090/2015 y en el art 7 del reglamento (UE) 536/2014 y considera que:

- El procedimiento para obtener el consentimiento informado (incluyendo las hojas de información al sujeto de ensayo y consentimientos informados mencionados en el encabezamiento), y el plan de reclutamiento de sujetos previsto son adecuados y cumplen con los requisitos para la obtención del consentimiento informado previstos en el capítulo II del Real Decreto 1090/2015.
- Las compensaciones previstas a los participantes son adecuadas, así como las previsiones de indemnización por daños y perjuicios que pueda sufrir el participante.
- El procedimiento previsto para el manejo de datos personales es adecuado.
- El uso futuro de las muestras biológicas obtenidas durante el ensayo se adecua a lo previsto en el Real Decreto 1716/2011.
- Para la realización del ensayo se consideran adecuados los centros e investigadores previstos en el anexo II a este dictamen, teniendo en cuenta las declaraciones de idoneidad emitidas por el promotor y por los responsables de las instituciones correspondientes.

Que este Comité decidió emitir DICTAMEN FAVORABLE en la reunión celebrada el día 19/12/16 (Acta nº 20.16).

Que en dicha reunión se cumplieron los requisitos establecidos en la legislación vigente –Real Decreto 1090/2015 – para que la decisión del citado CEIm sea válida.

Que el CEIm del Hospital Universitario Puerta de Hierro Majadahonda, tanto en su composición como en sus procedimientos, cumple con las normas de BPC (CPMP/ICH/135/95) y con la legislación vigente que regula su funcionamiento, y que la composición del CEIm del Hospital Universitario Puerta de Hierro Majadahonda es la indicada en el anexo I, teniendo en cuenta que en el caso de que algún miembro participe en el ensayo o declare algún conflicto de interés no habrá participado en la evaluación ni en el dictamen de la solicitud de autorización del ensayo clínico.

Lo que firmo en Majadahonda, a 19 de diciembre de 2016

 **Hospital Universitario
Puerta de Hierro
Majadahonda**
SaludMadrid
Comunidad de Madrid
**Comité Ético de Investigación
Clínica**

Fdo. Dra. Avendaño Solá

Anexo I

COMPOSICION DEL CEIm

Presidenta	Dra. Avendaño Solá (Farmacología Clínica)
Secretaria Técnica	Dra. Belén Ruiz Antorán (Farmacología Clínica)
Vocales	Dra. Fornet Ruiz (Anestesiología) Dr. García Marco (Hematología) Dr. Rodríguez Reina (Urología) (C. Ética Asistencial) Dr. Sánchez Ruiz (Oncología Médica) Dra. Rodríguez Marrodán (Farmacia Hospitalaria) Dra. Paloma Martín Acosta (Anatomía Patológica) (Comité Científico del IIS Puerta de Hierro) D ^a . R. Layunta (Enfermera) D. J. Moreno (Abogado) (Miembro ajeno al centro)

Anexo II

CENTROS E INVESTIGADORES PRINCIPALES PARTICIPANTES EN ESPAÑA

CÓDIGO: NADIM (GCEP 16/03)

NÚMERO EUDRACT: 2016-003732-20

TÍTULO: Ensayo clínico fase II, exploratorio y multicéntrico, de quimioinmunoterapia neo-adyuvante para el tratamiento del cáncer de pulmón no microcítico estadio III operable.

FECHA ACTUALIZACIÓN ANEXO II: 19/12/16

Investigador principal	Centro de realización del estudio
Dr. Mariano Provencio	Hospital Universitario Puerta de Hierro Majadahonda
Dra. Rosario García Campelo	Complejo Hospitalario Universitario A Coruña
Dr. Bartomeu Massuti	Hospital Universitario Alicante
Dra. Nuria Viñolas	Hospital Clínic de Barcelona
Dr. Álex Martínez Martí	Hospital Vall Hebrón
Dra. Margarita Majem	Hospital de la Santa Creu i Sant Pau
Dr. Santiago Viteri	Hospital Universitario Quirón-Dexeus / Instituto Oncológico Dr. Rosell
Dr. Guillermo López Vivanco	Hospital Universitario de Cruces
Dr. Isidoro C. Barneto Aranda	Hospital Universitario Reina Sofía
Dr. Ernest Nadal	ICO Hospitalet
Dr. Delvys Rodríguez	Hospital Universitario Insular de Gran Canaria
Dr. J.L. González Larriba	Hospital Clínico San Carlos
Dr. Manuel Dómine	Hospital Univer. Fundación Jiménez Díaz
Dr. José Miguel Sánchez	H.U. de la Princesa
Dr. Javier de Castro	Hospital Universitario la Paz
Dr. Manuel Cobo	Hospital Universitario Málaga General (Carlos Haya)
Dra. Raquel Marsé Fabregat	Hospital Universitario Son Espases
Dr. Ignacio Gil Bazo	Clínica Universidad de Navarra
Dr. David Vicente Baz	Hospital Universitario Virgen Macarena
Dr. Jesús Corral	Hospital Universitario Virgen del Rocío
Dra. Amelia Insa	Hospital Clínico Universitario de Valencia
Dr. Carlos Camps	Hospital General de Valencia
Dr. José Gómez Codina	Hospital Universitario y Politécnico La Fe
Dr. Joaquín Casal Rubio	Complejo Hospitalario Univer. De Vigo