



Universitat Oberta
de Catalunya

Efecto del tratamiento con Ibrutinib en pacientes de Leucemia Linfocítica Crónica: análisis de datos de scRNA-seq y ATAC-seq.

Javier Rey Barroso

Master universitario en Bioinformática y Bioestadística UOC/UB
Área 3, Subárea 5: Epigenómica y cáncer

Consultora: Izaskun Mallona González

Profesor responsable de la asignatura: Ferrán Prados Carrasco

24/06/2020



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-SinObraDerivada
[3.0 España de Creative Commons](#)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Efecto del tratamiento con Ibrutinib en pacientes de Leucemia Linfocítica Crónica: análisis de datos de scRNA-seq y ATAC-seq.</i>
Nombre del autor:	<i>Javier Rey Barroso</i>
Nombre del consultor/a:	<i>Izaskun Mallona González</i>
Nombre del PRA:	<i>Ferrán Prados Carrasco</i>
Fecha de entrega (mm/aaaa):	<i>06/2020</i>
Titulación::	<i>Master universitario en Bioinformática y Bioestadística UOC/UB</i>
Área del Trabajo Final:	<i>Área 3, Subárea 5: Epigenómica y cáncer</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Ibrutinib, scRNA-seq, ATAC-seq</i>

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

El presente trabajo estudia el efecto del tratamiento con Ibrutinib en pacientes de Leucemia Linfocítica Crónica (LLC) mediante el análisis transcriptómico (por scRNA-seq) y del estado de la cromatina (por ATAC-seq) de las células tumorales y su entorno inmune. La LLC se caracteriza por la acumulación masiva de células tumorales en órganos linfoides. Éstas son expulsadas a la sangre por efecto del tratamiento con un inhibidor de la Btk llamado Ibrutinib, el cual mejora fuertemente la salud de los pacientes.

Decidimos comprobar si el Ibrutinib afecta la maquinaria celular de adhesión y migración de células de LLC, T-CD4, T-CD8 y monocitos analizando datos de scRNA-seq y ATAC-seq procedentes de PBMC. Para ello utilizamos múltiples paquetes asociados a Bioconductor y R entre los que destacan los paquetes Seurat y GenomicRanges.

Los resultados de transcriptómica mostraron la reducción de la expresión de genes asociados a adhesión y migración como el complejo ARP2/3, RhoGTPasas e Integrinas en células tumorales y, en menor medida, monocitos.

Los resultados en el estudio de la apertura de la cromatina en promotores mostraron un silenciamiento en promotores asociados a múltiples vías de activación de células LLC, pero no en procesos de adhesión o migración celular.

Tras el estudio, podemos concluir que el tratamiento con Ibrutinib inhibe la expresión de la maquinaria de adhesión y migración de células LLC pero que este efecto parece no estar mediado por el silenciamiento de sus promotores. Estos resultados explican, en

parte, el fenómeno de éxodo de células LLC inducido por Ibrutinib.

Abstract (in English, 250 words or less):

We studied the effect of Ibrutinib treatment on Chronic Lymphocytic Leukaemia (CLL) by transcriptomic analyses(scRNA-seq) and chromatin landscape (ATAC-seq) of tumoral cells and their immune environment. CLL is characterized by the massive accumulation of tumoral cells in lymphoid organs. These cells exit to the blood as an effect of the Btk inhibitor Ibrutinib, which dramatically improves patients health.

We aimed to check if Ibrutinib impairs adhesion and migration of CLL cells, T-CD4, T-CD8 and monocytes by analizing scRNA-seq and ATAC-seq from PBMCs. For this purpose we used several R and Bioconductor packages, specially Seurat and GenomicRanges.

Transcriptomic results showed a reduction in adhesion and migration associated genes as ARP2/3 complex proteinsm RhoGTPases and Integrins, in tumoral cells and, to a lesser extent, monocytes.

Results in chromatin landscape analyses showed a silencing on promoters associated to several CLL activation pathways, but not in adhesion nor migration.

Thanks to this project, we can conclude that Ibrutinib treatment inhibits adhesion and migration machinery expression in CLL cells although this effect does not seem to be driven by promoter silencing.

These results may help to explain the CLL cell exodus from lymphoid organs in response to Ibrutinib

Índice

Lista de tablas.....	vi
Lista de figuras	vi
1. Introducción.....	7
1.1 Contexto y justificación del Trabajo.....	7
La leucemia linfocítica crónica.....	7
El microentorno tumoral.....	7
El tratamiento con Ibrutinib.....	7
La migración celular.....	9
Los receptores de quimioquinas.....	9
Las integrinas	9
El citoesqueleto de actina	10
Últimos resultados	10
El trabajo de Rendeiro y colaboradores.....	10
Las técnicas de transcriptómica y epigenómica utilizadas.	11
1.2 Objetivos del Trabajo.....	13
Objetivos generales:	13
Objetivos específicos:.....	13
1.4 Planificación del Trabajo	15
1.5 Breve sumario de productos obtenidos.....	19
1.6 Breve descripción de los otros capítulos de la memoria.....	20
2. Material y métodos	21
3. Efecto del Ibrutinib en la expresión génica de PBMC vía scRNA-seq	25
Lectura y evaluación de los datos.....	25
Anulación de los batch effects y selección de genes de interés	26
Análisis de componentes principales (PCA) y elección de dimensiones	27
<i>Clustering</i> de poblaciones.....	30
División de las muestras en antes y durante el tratamiento e identificación de las poblaciones celulares	31
Análisis de expresión diferencial por ibrutinib y Gene Ontology.....	34
Células tumorales	34
Linfocitos T y monocitos	37
4. Efecto del Ibrutinib en la apertura de la cromatina de células immunes vía ATAC- seq. 43	
Efecto en células tumorales (CLL).....	43
Lectura y carga datos	43
Filtrado de picos detectados	43
Anotación de picos	43
Estudio diferencial del efecto del ibrutinib. Búsqueda de picos comunes.....	44
Análisis de componentes principales.....	46
Definición del grupo de picos consenso asociados a promotores	47
Establecimiento de un índice de apertura de la cromatina en promotores con respecto al tiempo de tratamiento.....	47
Detección de promotores afectados por el tratamiento con ibrutinib y evolución temporal.....	47
Estudio GO de los genes con promotores diferencialmente accesibles.....	48

Efecto en linfocitos T y en monocitos	50
5. Conclusiones.....	51
6. Glosario	54
7. Bibliografía.....	55
8. Anexos	58

Lista de tablas

Tabla I: muestras ATAC-seq.

Tabla II: muestras scRNA-seq.

Tabla III. Principales marcadores de los *clusters* detectados.

Tabla IV: Términos GO destacados por PANTHER en la inhibición de promotores por efecto del Ibrutinib.

Lista de figuras

Figura 1: Rutas de señalización afectadas por Ibrutinib

Figura 2: Migración de células B durante su desarrollo y función

Figura 3: Gráficas de evaluación de la calidad de las muestras scRNA-seq utilizadas.

Figura 4: Correlación entre índices de calidad de las muestras.

Figura 5. Genes de interés.

Figura 6: Distribución de las células de cada muestras en las primeras dos componentes y genes con mayor peso en la distribución.

Figura 7: Expresión de los principales genes de las primeras 10 componentes.

Figura 8. Elbow plot

Figura 9. Distribución de las poblaciones detectadas; representación por PCA.

Figura 10. Distribución de las poblaciones detectadas; representación por UMAP.

Figura 11: Expresión de marcadores en los diferentes clusters y efecto del Ibrutinib

Figura 12: Expresión relativa de marcadores; UMAP.

Figura 13: Asignación de poblaciones a los clusters detectados; UMAP.

Figura 14: Asignación de poblaciones, UMAP muestra a muestra.

Figura 15: Genes con expresión diferencial por efecto del tratamiento con ibrutinib en células tumorales.

Figura 16. Análisis GO con EnrichR en células tumorales.

Figura 17. Efecto del Ibrutinib en la expresión relativa de genes de adhesión y migración de células tumorales.

Figura 18: Genes con expresión diferencial por efecto del tratamiento con ibrutinib en linfocitos T CD8

Figura 19: Efecto del Ibrutinib en la expresión relativa de genes de interés en linfocitos T CD8.

Figura 20 Genes con expresión diferencial por efecto del tratamiento con ibrutinib en linfocitos T CD4.

Figura 21 Genes con expresión diferencial por efecto del tratamiento con ibrutinib en monocitos.

Figura 22: Principales términos GO y genes asociados, inhibidos por Ibrutinib en monocitos.

Figura 23: Resumen de la anotación de picos.

Figura 24. Coincidencia de picos entre muestras de CLL.

Figura 25. Coincidencia de picos entre muestras de CLL. Upset plots

Figura 26: Componentes principales.

Figura 27: Cambios en la apertura de promotores de células LLC por efecto del tratamiento con Ibrutinib

1. Introducción

1.1 Contexto y justificación del Trabajo

La leucemia linfocítica crónica.

La Leucemia Linfocítica Crónica (LLC) es una enfermedad caracterizada por la proliferación clonal de linfocitos B maduros y su acumulación masiva en sangre, médula ósea, bazo y nódulos linfáticos (Montserrat & Rozman, 1995). Se trata de la leucemia más común en países occidentales (4 casos cada 100000 habitantes en EEUU) y se caracteriza por su aparición en personas mayores (con una media de edad del paciente de 72 años) y una evolución lenta (Hallek, 2019). Un estudio reciente basado en el estudio de largas cohortes identificó como principales orígenes de la enfermedad la pérdida o ganancia de grandes cantidades de material cromosómico (por ejemplo, la delección del brazo 13q, la delección del 11q o una trisomía del cromosoma 12), las cuales se ven seguidas de mutaciones adicionales que vuelven a la leucemia más agresiva (Landau et al., 2015). Algunos de los genes mutados de forma recurrente son NOTCH1 , MYD88 , TP53 , ATM , SF3B1 , FBXW7 , POT1 , CHD2 , RPS15 , IKZF3 , ZNF292 , ZMYM3 , ARID1A y PTPN11, siendo las vías de procesado de ARN, la actividad MYC y la señalización por MAP quinasas las consideradas centrales en el desarrollo de la LLC (Landau et al., 2015).

El microentorno tumoral

A pesar de que el diagnóstico de la enfermedad se produce a partir de la acumulación continuada de linfocitos B en sangre periférica (≥ 5000 linfocitos B/ μL , durante más de tres meses) (Hallek, 2019), se considera que el origen y desarrollo de la enfermedad se encuentra en los tejidos. En los nódulos linfáticos, el bazo y la médula espinal, las células tumorales encuentran e inducen un microentorno que permite y estimula su crecimiento(Arruga et al., 2020). Concretamente, el entorno inmune, compuesto por macrófagos, linfocitos T CD4 y CD8 y células dendríticas, entre otros tipos celulares, les aporta señales de supervivencia, activación y proliferación. Esto ocurre a través de citoquinas, quimioquinas y factores angiogénicos, los cuales interaccionan con receptores de superficie y proteínas de adhesión de las células LLC (Burger et al., 2009). Por otra parte, la interacción de las células tumorales con las células inmunes también altera el comportamiento de estas últimas, fomentando la inhibición de células CD8 y CD4 efectoras, así como la diferenciación de macrófagos en tipo M2, antiinflamatorios, lo que genera una tolerancia inmune al tumor y, por tanto, permite su expansión (Arruga et al., 2020). La importancia de estas interacciones es tal, que se ha visto que parte del efecto clínico de terapias diseñadas para atacar específicamente las células tumorales se debe a efectos indirectos de los fármacos en células del microentorno inmune, como es el caso de Idelalisib o Ibrutinib (Maffei et al., 2015).

El tratamiento con Ibrutinib

El ibrutinib es un inhibidor dirigido contra la quinasa BTK (Bruton's tyrosine kinase) utilizado actualmente como uno de los fármacos más efectivo contra la LLC. Funciona

mediante su unión irreversible al dominio quinasa de la BTK, proteína implicada en la transducción de gran parte de las señales derivadas de la activación del BCR (B cell receptor), así como de la recepción de quimioquinas (Hallek, 2019).

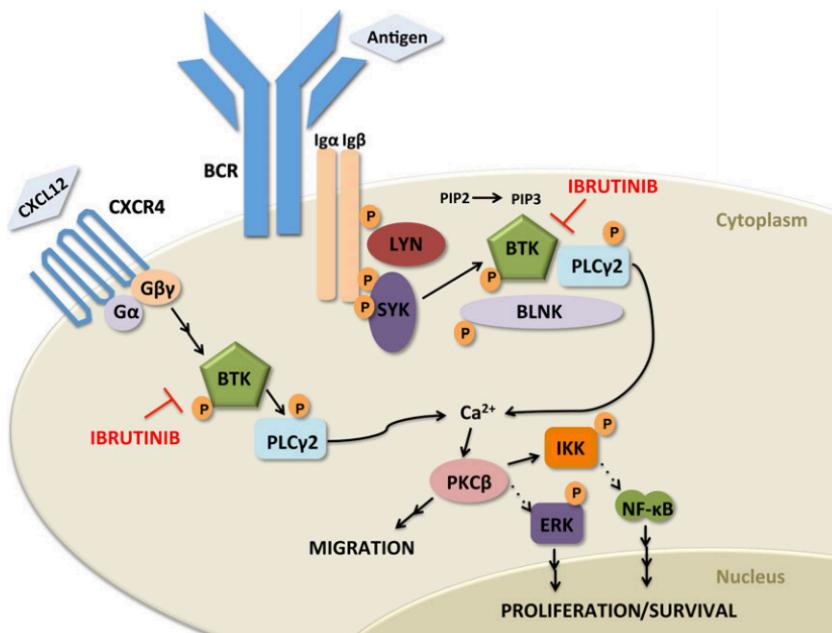


Figura 1: Rutas de señalización afectadas por Ibrutinib (Sivina et al., 2014).

El principal efecto observado en pacientes tratados con Ibrutinib es la salida masiva de células tumorales de los ganglios y el bazo hacia la sangre, donde estas células adquieren un fenotipo quiescente y finalmente mueren por ausencia de señales pro-supervivencia (Chang et al., 2013). Estudios llevados a cabo en pacientes muestran que el éxodo de los ganglios linfáticos se correlaciona con una menor habilidad de las células tumorales para migrar en respuesta a quimioquinas como CXCL12 (asociada a la entrada en la médula ósea) (Ponader et al., 2012) y una menor capacidad de adhesión a fibronectina (Herman et al., 2015). Es por ello que el estudio del efecto de tratamientos contra la LLC como el Ibrutinib en la maquinaria celular dedicada a la adhesión y migración resulta clave en la comprensión de su modo de acción, así como de sus posibles efectos secundarios.

Por otra parte, se ha demostrado que el Ibrutinib, a pesar de estar diseñado contra una quinasa específica de células B, también afecta de forma directa e indirecta al resto de poblaciones inmunes que conforman el microentorno tumoral de la LLC (Maffei et al., 2015). Destaca la inhibición de la quinasa ITK en linfocitos T, lo que induce un cambio en la diferenciación de los mismos hacia un perfil Th1 y reactiva la respuesta inmune contra las células tumorales (Dubovsky et al., 2013). También se han descrito efectos del Ibrutinib en el desarrollo de macrófagos a partir de monocitos (Ferrarini et al., 2019). Aparte de estos efectos directos, existen crecientes pruebas de que el Ibrutinib alteraría los contactos entre las células tumorales y el resto de poblaciones inmunes, lo que disminuiría fuertemente la recepción de señales de supervivencia y proliferación, provocando la reducción y posible remisión del tumor (Maffei et al., 2015).

La migración celular

Una de las habilidades compartidas por las células inmunes, así como vitales para su desarrollo y función, es la capacidad de viajar entre los distintos tejidos y la sangre. En el caso de las células B, éstas son originadas en la médula ósea a partir de precursores hematopoyéticos, tras lo cual entran en el torrente circulatorio y, desde este, colonizan los órganos linfoides secundarios (ganglios linfáticos, bazo ...). En caso de respuesta inflamatoria, los linfocitos B viajarán hasta el tejido afectado para ejercer su función como células B plasmáticas, mediante la secreción de anticuerpos. Potencialmente, parte de estas células se convertirán en células B memoria y regresarán a los órganos linfoides (Imagen 1) (Otagiri et al., 2017)

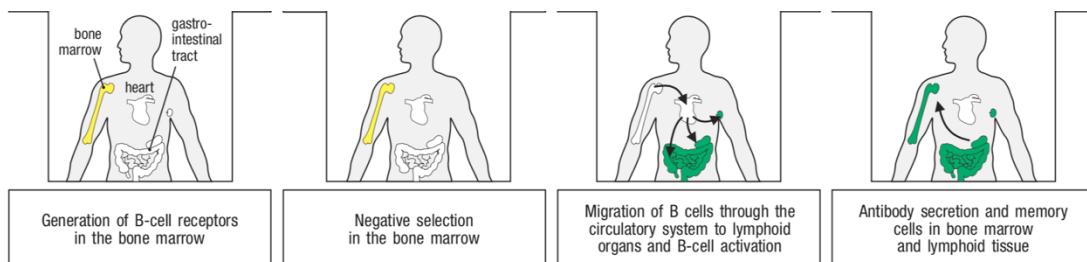


Figura 2: Migración de células B durante su desarrollo y función(Otagiri et al., 2017).

Esta capacidad única de dirigirse y colonizar entornos tan diferentes se debe, principalmente, a la integración de tres familias de proteínas: Los receptores de quimioquinas, las integrinas y el citoesqueleto de actina.

Los receptores de quimioquinas

Las quimioquinas son un amplio grupo de citoquinas que median el control del tráfico de células inmunes en el organismo. Se trata de pequeñas moléculas capaces de inducir migración celular en aquellas células con los receptores adecuados. Se han descrito 19 receptores de quimioquinas humanos. Se trata de proteínas de membrana pertenecientes a la clase “G protein-coupled receptors (GPCR)”, capaces de reconocer de manera específica las quimioquinas e inducir una cascada de señalización intracelular que dirige la migración celular (Rossi & Zlotnik, 2000). Cada quimioquina es reconocida por uno o varios receptores y viceversa y se ha observado que la activación de un mismo receptor por dos quimioquinas diferentes puede inducir rutas de señalización (y respuestas) diferentes(Raju et al., 2015).

Las integrinas

Las integrinas componen una familia de receptores de membrana heterodímericos, los cuales son los principales responsables de la adhesión de los leucocitos al entorno extracelular. Se componen de una cadena alfa y una beta, las cuales van a conformar un largo dominio extracelular, responsable de la unión a proteínas de la matriz extracelular (colágenos, lamininas, fibronectina ...) así como ligandos presentes en otras células (ICMs, VCAMs...). En su parte citoplasmática, ambas cadenas forman un pequeño

dominio implicado en la regulación del estado de activación, transducción de la señal y la unión al citoesqueleto. Esta familia de proteínas está sometida a una fuerte regulación, no sólo a nivel de expresión sino también en la adopción de un estado conformacional activo o inactivo. (Fagerholm et al., 2019)

El citoesqueleto de actina

Las actinas conforman una familia de proteínas globulares con la capacidad de polimerizar en una gran variedad de formas filamentosas, cuyo empaquetado y ramificación conduce a la generación de una gran variedad de estructuras que permiten la integración e las señales extracelulares y el remodelado de la forma celular en respuesta a las mismas, lo que permite en última instancia la adhesión y migración celulares. La conexión entre integrinas y citoesqueleto de actina va a permitir a la célula sentir y reaccionar a los cambios fisicoquímicos del entorno, lo que se conoce como “mecanosensibilidad” y “mecanotraducción”. Por otra parte, la estimulación de la polimerización orientada de la actina en respuesta a la recepción de quimioquinas permite la direccionalidad de la migración (Dupré et al., 2015). Estos son sólo dos ejemplos del alto grado de interacción el citoesqueleto de actina con el resto de componentes celulares, el cual requiere un control espacio-temporal preciso, ejercido por una gran variedad de proteínas conocidas como reguladores de actina.

Últimos resultados

A la vista de la importancia de la adhesión y migración celular en el desarrollo de la LLC, así como de su implicación en los efectos beneficiosos y secundarios del tratamiento con Ibrutinib, el proyecto COMPUTREAT-CLL seleccionó una cohorte de más de 50 pacientes para el estudio de sus efectos. Este se realizó mediante el análisis simultáneo, entre otros, de la presencia celular en los diferentes órganos del paciente y la expresión de receptores de quimioquinas e integrinas de las células inmunes de sangre periférica, así como su capacidad de migración *in vitro* frente a diferentes quimioquinas. Responsable de los ensayos de citometría y migración, nuestro grupo pudo demostrar que el tratamiento inducía una reducción de la expresión de numerosos receptores de quimioquinas, con una consecuente inhibición de la capacidad de migración de las células tumorales frente a CXCL12, CXCL13 y CCL19. In vitro, la presencia de Ibrutinib alteró la capacidad de adhesión de las células, así como su velocidad de movimiento en ausencia de quimioquina. Además, estos efectos fueron compartidos por las células T. Estos resultados, aún no publicados, indican que la inhibición por Ibrutinib sobrepasaría los efectos asociados a la recepción de quimioquina, y abren la puerta a la exploración de otras rutas de señalización afectadas, como la regulación del citoesqueleto de actina, clave en los procesos afectados.

El trabajo de Rendeiro y colaboradores

De manera independiente y complementaria, acaba de publicarse un artículo titulado “Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL” (Rendeiro et al., 2020). En este, los autores analizaron la evolución temporal de una pequeña cohorte de 7 pacientes tratados con Ibrutinib, mediante fenotipado por citometría, análisis transcriptómico de célula única (scRNA-seq) y mapeado de la cromatina (ATAC-seq) en seis poblaciones inmunes diferentes. La

integración de los resultados de estas tres técnicas permitió analizar los efectos consenso del Ibrutinib así como la variabilidad entre pacientes.

El primer efecto común detectado en células de LLC fue el rápido descenso de la unión de NF-kB a sus dianas, seguido de una reducción de la actividad de factores de transcripción ligados a la diferenciación celular, una erosión de la “identidad celular” LLC y la adopción de un programa génico cercano a la quiescencia. En cuanto a la variabilidad entre pacientes, se comprobó que, aunque todos los pacientes siguieron prácticamente el mismo proceso, la velocidad de éste difería, lo que se correlacionaba con el grado global de respuesta del paciente al tratamiento con Ibrutinib.

En cuanto a los efectos en el microentorno inmune, se observó que los principales cambios observados en células tumorales también sucedieron en el resto de poblaciones, especialmente en células B no tumorales y T. De manera específica, se observó un viraje de los monocitos hacia un fenotipo tipo M1 (pro-inflamatorio), así como una regulación de la cromatina diferencial entre linfocitos T CD4 y CD8.

Curiosamente, el análisis de los genes alterados por el tratamiento con Ibrutinib destacó genes asociados a adhesión y migración como, por ejemplo, CD44, CD99, CD53, ICAM3 y ITGB7 en el ATAC-seq, así como el enriquecimiento de términos GO (“Gene Ontology”) como “CDC42 signaling”, “Rac1 signaling”, “CXCR4-signaling”, “Integrin-mediated adhesion” o “Focal Adhesion” en el caso del scRNA-seq (Supplementary Figure 5,). Estos resultados fueron brevemente discutidos por los autores, quienes los englobaron dentro de la pérdida de identidad celular provocada por el Ibrutinib.

Lo complementario de los resultados obtenidos por el grupo de Christoph Bock y los obtenidos por el nuestro, llevan a pensar que un análisis detallado de los datos generados en el estudio, centrado esta vez en el efecto el Ibrutinib en genes asociados a adhesión y migración celular, podría desvelar nuevas rutas de señalización y genes diana afectados por el tratamiento, los cuales ayudarían a explicar los efectos celulares observados *in vitro* y *ex vivo*, así como ayudar comprender la respuesta de cada paciente al tratamiento.

Las técnicas de transcriptómica y epigenómica utilizadas.

A fin de comprender el potencial y limitaciones de las dos técnicas de secuenciación masiva utilizadas en el desarrollo del proyecto, a continuación e incluye una breve explicación de cada una de ellas.

scRNA-seq

Se trata de una técnica derivada del RNA-seq, un protocolo que permite la cuantificación del transcriptoma de un conjunto de células, así como su comparación entre diferentes muestras. A diferencia de su predecesor, cuyos valores corresponden al valor medio de las miles de células que componen cada muestra, el scRNA-seq permite analizar el transcriptoma de cada célula de manera individual. Este avance permite, entre otras ventajas, la identificación y estudio independiente de los diferentes tipos celulares que componen muestras heterogéneas, además de permitir la observación de la heterogeneidad de los patrones de expresión dentro de un mismo tipo celular. El

aumento de la resolución en el estudio de transcriptomas debido a la llegada del scRNA-seq, ha supuesto también un reto para el manejo, análisis e interpretabilidad de los datos, debido principalmente al aumento del ruido de fondo asociado a la técnica y al alto número de dimensiones de los resultados (Olsen & Baryawno, 2018).

Los autores utilizaron la tecnología 10x Genomics' single-cell RNA-seq (scRNA-seq) Chromium™. Brevemente, un sistema de microfluídica permite generar cesículas llamadas GEMs (Gel Beads in Emulsion) las cuales contienen una célula, los reactivos para la reacción de retrotranscripción y oligonucleótidos de identificación (barcoding). La generación de miles de GEMs permite generar cDNAs marcados correspondientes a una única células. Con estos, se realizan las librerías de cDNA, las cuales son secuenciadas en máquinas Illumina. Tras ello, las lecturas generadas son asociadas a su célula de origen y cuantificadas gracias al software Cell Ranger (10xGenomics.com)

ATAC-seq

Esta técnica, llamada “Assay of Transposase Accessible Chromatin sequencing”, publicada originalmente en 2013, es una de las más utilizadas actualmente en el estudio del paisaje epigenético de poblaciones celulares. Concretamente, el ensayo se basa en la utilización de una transposasa, llamada Tn5, genéticamente modificada que sistemáticamente corta la cromatina abierta, dejando extremos de 9 pares de bases, ligando en esta región adaptadores de secuenciación. En el proceso, los extremos son reparados, generan una duplicación de 9 pares de bases. Esto será utilizado tras la secuenciación para el alineamiento de las regiones abiertas detectadas (Yan et al., 2020). La técnica requiere una cantidad muy baja de células y posee una alta sensibilidad y especificidad. Lo reciente de la técnica hace que aún no existan protocolos de análisis estandarizados. A pesar de ello, esta técnica ha permitido el análisis de la apertura diferencial e la cromatina entre el proceso hematopoyesis y el desarrollo de leucemia (Corces et al., 2016) o establecer el mapa del estado de la cromatina en el proyecto TGCA (The Cancer Genome Atlas).

1.2 Objetivos del Trabajo

Objetivos generales:

- Analizar el efecto del Ibrutinib en el control epigenético y la expresión de genes relacionados con la adhesión y migración de células tumorales e inmunes en pacientes de LLC.
- Evaluar la importancia de estos cambios en la respuesta del paciente al tratamiento con Ibrutinib.

Objetivos específicos:

- Analizar el efecto del Ibrutinib en el control epigenético y la expresión de genes relacionados con la adhesión y migración de células tumorales e inmunes en pacientes de LLC.
- Trazar la evolución temporal de la expresión de genes asociados a la adhesión y migración celulares en células tumorales de LLC en respuesta a ibrutinib mediante el análisis de datos de single-cell RNA-seq en diferentes puntos del tratamiento.
- Analizar el estado de la cromatina en las áreas de expresión de genes asociados a adhesión y migración celular células tumorales de LLC en y su evolución en respuesta a Ibrutinib mediante el análisis de datos de ATAC-seq.
- Evaluar la heterogeneidad poblacional en células tumorales de LLC, así como la evolución de la misma en respuesta a Ibrutinib. Análisis mediante clustering no supervisado y correlación con genes de colonización.
- Estimar los efectos colaterales del ibrutinib en el perfil de adhesión y migración otras poblaciones celulares (principalmente linfocitos T). Análisis de datos de sc-RNA-seq y ATAC-seq
- Analizar la correlación entre los cambios de perfil de adhesión y migración de células tumorales de LLC detectados por scRNA-seq y ATAc-seq y la evolución de la linfocitosis durante el tratamiento con Ibrutinib.
- Estimar la correlación entre los efectos en otras poblaciones celulares y la evolución de la linfocitosis durante el tratamiento con Ibrutinib.

1.3 Enfoque y método seguido

Desde hace décadas, el estudio del efecto de diferentes fármacos anti-cáncer, así como de sus efectos colaterales, ha revelado la complejidad del modo de acción de muchos de ellos. Tradicionalmente, estas investigaciones partían del establecimiento de hipótesis sobre un número limitado dianas probables del fármaco.

En los últimos años, la aparición de las tecnologías de transcriptómica, especialmente las de nueva generación, permiten orientar la investigación de una manera global y, en la mayor parte de los casos, no dirigida. Este nuevo enfoque permite descubrir las principales vías de señalización implicadas en la acción del fármaco, así como dianas inesperadas. Sin embargo, estos estudios generan tal cantidad de información que una gran parte de la misma queda a menudo sin explorar y, en ocasiones, la interpretación “global” de los resultados resulta altamente compleja.

El presente estudio ha consistido en la explotación de los resultados de una serie de experimentos de análisis transcriptómico y de apertura de la cromatina realizado en pacientes de LLC tanto antes, como a diferentes puntos de su tratamiento con Ibrutinib. La originalidad del mismo reside en que se ha enfocado en el estudio de genes de adhesión y migración, estableciendo la hipótesis de que el Ibrutinib induce cambios en la expresión de éstos que podrían explicar la respuesta global al tratamiento, así como las diferencias de la efectividad del mismo entre pacientes.

La combinación de la investigación ómica y la búsqueda individual de rutas afectadas podría explicar el fuerte aumento de células tumorales circulantes asociado al tratamiento, así como la aparición de resistencias al mismo.

Por otra parte, este fármaco, dirigido contra la kinasa Btk (específica de células B) es conocido por sus efectos colaterales en otras células inmunes. Dado que son estas últimas las encargadas de favorecer o combatir el desarrollo tumoral, resulta clave el análisis paralelo de los efectos que el Ibrutinib puede tener en la población tumoral y en el entorno inmune. Este punto queda perfectamente cubierto gracias a la utilización de *single cell* RNA-seq en muestras de sangre periférica, lo que permite diseccionar cuál es el de respuesta para cada tipo celular. Por otra parte, el análisis por ATAC-seq de las principales poblaciones inmunes de la sangre, esta vez aisladas previamente mediante citometría, permite obtener respuestas específicas de tipo celular evitando una escasez excesiva de datos.

Se trabajó a partir de los datos procesados y alineados por los autores del artículo. Para abordar el proyecto se utilizarán diferentes paquetes de análisis basados en R (R Core team & Computing., 2020), los cuales son ampliamente utilizados y validados por la comunidad científica actual.

El procedimiento, resumido, del análisis a realizar tanto para el scRNA como para el ATAC-seq es el siguiente:

- a) Procesado de control de calidad
- b) Selección de variables de interés
- c) Análisis de componentes principales

- d) Detección y corrección de batch effects
- e) Clustering no supervisado de poblaciones (Louvain)
- f) Representación vía reducción de dimensionalidad (UMAP)
- g) Detección de principales marcadores asociados al tratamiento con Ibrutinib
- h) Correlación entre estos y términos asociados a adhesión y migración.

1.4 Planificación del Trabajo

Tareas:

1. Exploración de paquetes a utilizar en los diferentes pasos

2.2.1. Analizar el efecto del Ibrutinib en el control epigenético y la expresión de genes relacionados con la adhesión y migración de células tumorales e inmunes en pacientes de LLC.

2.2.1.1. Trazar la evolución temporal de la expresión de genes asociados a la adhesión y migración celulares en células tumorales de LLC en respuesta a ibrutinib mediante el análisis de datos de single-cell RNA-seq en diferentes puntos del tratamiento.

- 2.2.1.1.1. Procesado de control de calidad (Seurat)
- 2.2.1.1.2. Selección de variables de interés (Seurat)
- 2.2.1.1.3. Asignación al tipo celular en función de marcadores CD5 y CD19
- 2.2.1.1.4. Análisis de componentes principales (Seurat)
- 2.2.1.1.5. Detección y corrección de batch effects
- 2.2.1.1.6. Detección de principales marcadores de las subpoblaciones (Enrichr API)
- 2.2.1.1.7. Correlación entre estos y nuestros marcadores de interés. (GO Adhesion, chemotaxis, integrin etc.).

2.2.1.2. Analizar el estado de la cromatina en las áreas de expresión de genes asociados a adhesión y migración celular células tumorales de LLC en y su evolución en respuesta a Ibrutinib mediante el análisis de datos de ATAC-seq.

- 2.2.1.2.1. Procesado de control de calidad (GenomicRanges)
- 2.2.1.2.2. Selección de picos de interés (ChiPseeker)
- 2.2.1.2.3. Análisis de componentes principales
- 2.2.1.2.4. Detección y corrección de batch effects (Combat)
- 2.2.1.2.5. Análisis diferencial de presencia de picos por efecto del Ibrutinib
- 2.2.1.2.6. Correlación entre estos y nuestros marcadores de interés. (GO Adhesion, chemotaxis, integrin etc.).

2.2.1.3. Evaluar la heterogeneidad poblacional en células tumorales de LLC, así como la evolución de la misma en respuesta a Ibrutinib. Análisis mediante clustering no supervisado y correlación con genes de colonización.

- 2.2.1.3.1. Clustering no supervisado de poblaciones (Louvain, Seurat)
- 2.2.1.3.2. Representación vía reducción de dimensionalidad (tSNE, Seurat)

2.2.1.3.3. Correlación entre estos y nuestros marcadores de interés. (GO Adhesion, chemotaxis, integrin etc., así como lista personal de genes de interés).

2.2.1.4. Estimar los efectos colaterales del ibrutinib en el perfil de adhesión y migración otras poblaciones celulares (principalmente linfocitos T). Análisis de datos de sc-RNA-seq y ATAC-seq

Sc-RNA-seq

2.2.1.4.1. Asignación al tipo celular en función de marcadores específicos de tipo celular

2.2.1.4.2. Análisis de componentes principales (Seurat)

2.2.1.4.3. Detección y corrección de batch effects

2.2.1.4.4. Detección de principales marcadores asociados a tratamiento (Enrichr API)

2.2.1.4.5. Correlación entre estos y nuestros marcadores de interés. (GO Adhesion, chemotaxis, integrin etc.).

ATAC-seq

Para cada tipo celular:

2.2.1.2.7. Procesado de control de calidad (GenomicRanges)

2.2.1.2.8. Selección de picos de interés (ChiPseeker)

2.2.1.2.9. Análisis de componentes principales

2.2.1.2.10. Detección y corrección de batch effects (Combat)

2.2.1.2.11. Análisis diferencial de presencia de picos por efecto del Ibrutinib

2.2.1.2.12. Correlación entre estos y nuestros marcadores de interés. (GO Adhesion, chemotaxis, integrin etc.).

2.2.2. Evaluar la importancia de estos cambios en la respuesta del paciente al tratamiento con Ibrutinib.

2.2.2.1. Analizar la correlación entre los cambios de perfil de adhesión y migración de células tumorales de LLC detectados por scRNA-seq y ATAC-seq y la evolución de la linfocitosis durante el tratamiento con Ibrutinib.

2.2.2.1.1. Representación de la curva de linfocitosis en respuesta a Ibrutinib de cada paciente

2.2.2.1.2. División en grupos: Alta, media y baja linfocitosis.

2.2.2.1.3. Correlación de los cambios de expresión de interés detectados y la clasificación por linfocitosis de los pacientes.

2.2.2.2. Estimar la correlación entre los efectos en otras poblaciones celulares y la evolución de la linfocitosis durante el tratamiento con Ibrutinib.

2.2.2.2.1. Correlación de los cambios de expresión de interés detectados para cada tipo celular y la clasificación por linfocitosis de los pacientes

Calendario:

Marzo de 2020						
LUN 24	MAR 25	MIÉ 26	JUE 27	VIE 28	SÁB 29	DOM 1 de mar
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
1. Exploración de paquetes						
2.2.1.1 y 4 scRNA-seq LLC y células inmunes						
2.2.1.1/4, 1 a 4, preparación de datos						
23	24	25	26	27	28	29
2.2.1.1 y 4 scRNA-seq LLC y células inmunes						
2.2.1.1/4, 1 a 4, preparación de datos						
30	31	1 de abr	2	3	4	5
2.2.1.1 y 4 scRNA-seq LLC y células inmunes						
2.2.1.1/4, 6 y 7, detección de marcadores						

Introducción

Abril de 2020

Hoy	<	>	Abril de 2020	Search	?	⚙️	Mes	UOC	☰
LUN	MAR	MIÉ	JUE	VIE	SÁB	DOM			
30	31	1 de abr	2	3	4	5			
	2.2.1.1 y 4 scRNA-seq LLC y células inmunes								
	2.2.1.1/4 6 y 7. detección de marcadores								
6	7	8	9	10	11	12			
	2.2.1.3 Heterogeneidad en LCC								
	2.2.1.3. 1 a 3 Clustering y correlación con genes de interés								
13	14	15	16	17	18	19			
	2.2.1.2 y 4 ATAC-seq LLC y células inmunes								
	2.2.1.2/4 1 a 3. Procesado de datos			2.2.1.2/4.4 Batch effects					
20	21	22	23	24	25	26			
	2.2.1.2 y 4 ATAC-seq LLC y células inmunes								
	2.2.1.2/4.4 Batch effects			2.2.1.2/4.5 Cicero					
27	28	29	30	1 de may	2	3			
	2.2.1.2 y 4 ATAC-seq LLC y células inmunes								
	2.2.1.2/4.5 Cicero		2.2.1.2/4 6 y 7. Detección de marcadores						

Mayo de 2020

Hoy	<	>	Mayo de 2020	Search	?	⚙️	Mes	UOC	☰
LUN	MAR	MIÉ	JUE	VIE	SÁB	DOM			
27	28	29	30	1 de may	2	3			
	2.2.1.2 y 4 ATAC-seq LLC y células inmunes								
	2.2.1.2/4.5 Cicero		2.2.1.2/4 6 y 7. Detección de marcadores						
4	5	6	7	8	9	10			
	2.2.2.1. Correlación LLC y linfocitosis		2.2.2.2. Correlación linfocitosis y entorno inmune						
	2.2.2.1.1 Análisis de linfocitosis		Correlación con scRNA-seq LLC	2.2.2.2.1 Correlación con linfocitosis					
11	12	13	14	15	16	17			
	3. Puesta en formato de resultados								
18	19	20	21	22	23	24			

Hitos:

- scRNA-seq en células LLC y entorno immune:
 - Lectura y preparación de los datos completada
 - Batch effects eliminados
 - Marcadores detectados, comprobación de nuestros genes de interés realizada
- Heterogeneidad poblacional de LLC:
 - Clustering no supervisado realizado y visualización disponible
 - Detección de cambios en la proporción de las subpoblaciones observadas
 - Asociación a perfiles de expresión de interés realizada
- ATAC-seq en células LLC y entorno immune:
 - Lectura y preparación de los datos completada
 - Batch effects eliminados
 - Picos consenso establecidos
 - Marcadores detectados, comprobación de nuestros genes de interés realizada.
- Correlación con linfocitos:
 - Correlación con resultados en LLC realizada
 - Correlación con resultados de entorno inmune realizada

5. Resultados esperados:

Se concretan en cuatro entregables:

- 5.1. Plan de trabajo
- 5.2. Memoria
- 5.4. Presentación virtual
- 5.5. Autoevaluación del proyecto

6. Estructuración del proyecto:

EL presente proyecto e estructura en 6 fases:

- Definición de los contenidos del trabajo
- Establecimiento de un plan de trabajo
- Desarrollo del trabajo
- Cierre de la memoria
- Elaboración de la presentación
- Defensa pública

1.5 Breve sumario de productos obtenidos

- Memoria del proyecto
- Presentación virtual
- Código para la generación de los resultados
- Tablas anexas con resultados completos

1.6 Breve descripción de los otros capítulos de la memoria

2. Material y métodos. Descripción del origen y formato de los datos utilizados en el proyecto. Descripción de los métodos utilizados para su análisis.
3. Efecto del Ibrutinib en la expresión génica de PBMC vía scRNA-seq. Estudio de los datos de transcriptómica
4. Efecto del Ibrutinib en la apertura de la cromatina de células immunes vía ATAC-seq. Estudio de los datos de apertura de la cromatina.
5. Conclusiones: Discusión de los resultados y perspectivas.

2. Material y métodos

Datos analizados

Para la realización del estudio, partimos de los datos procesados publicados en el artículo de Rendeiro y colaboradores, los cuales conforman la Superserie GSE111015, en el portal GEO (Gene Expression Omnibus). Esta se compone de:

- 176 muestras ATAC-seq procedentes de células inmunes (aisladas de pacientes de LLC a diferentes tiempos de tratamiento (GSE111013).
- 12 muestras scRNA-seq procedentes de células mononucleares de sangre periférica (“PBMC”) de pacientes de LLC a diferentes tiempos de tratamiento (GSE111014).

A continuación se muestran dos tablas que resumen las muestras disponibles:

ATAC-seq									
Patient	Time point (days)								
CLL1	0			3		30	120		
CLL2	0			3	8	30	120		
CLL4	0		2	3					
CLL5	0	1	2	3	8	30		150	240
CLL6	0					30	120		240
CLL7	0	1		3	8	30		150	
CLL8	0			3		30			

Tabla I: muestras ATAC-seq. Para cada punto, se realizaron muestras de células B- CLL, B no tumorales, T-CD4, T-CD8, Monocitos y células NK.

sc-RNaseq				
Patient	Time point (days)			
CLL8	0	30	120	
CLL6	0	30	120	280
CLL5	0	30	150	
CLL1	0	30		

Tabla II: muestras scRNA-seq. Muestras generadas a partir de PBMCs.

Métodos

En ambos casos se utilizó R (R Core team & Computing., 2020) como programa principal de tratamiento de datos, apoyado por el interfaz RStudio (Allaire, 2015) y por el conjunto de herramientas que conforman Bioconductor (Gentleman et al., 2004). Los informes fueron realizados utilizando Rmarkdown (Sauer, 2019).

scRNA-seq

El análisis de datos de transcriptómica se realizó en su mayor parte gracias al paquete **Seurat**(Stuart et al., 2018), desarrollado específicamente para el análisis de datos scRNA-seq así como su integración con otras tecnologías single cell. Los autores pusieron especial énfasis en que el programa sea abordable para usuarios sin experiencia, añadiendo diferentes “vignettes” o tutoriales que permiten una mejor comprensión de su funcionamiento. EN el presente trabajo se siguieron principalmente dos de esos tutoriales: “Guided tutorial---2,700 PBMCs” y “Stimulated vs Control PBMCs”.

Otros paquetes requeridos para la ejecución del análisis:

- dplyr
- Seurat
- patchwork
- umap-learn (python)
- enrichR
- topGO
- limma
- GO.db
- org.Hs.eg.db
- mnormt
- metap

ATAC-seq

Para analizar los datos de apertura de la cromatina se siguió, al menos parcialmente, el tutorial “Analysis of ATAC-seq data in R and Bioconductor” de Lewis Carroll, basado en el uso de GenomicRanges (Lawrence et al., 2013), aunque se realizaron modificaciones sustanciales al protocolo detallado en el mismo.

Otros paquetes requeridos para la ejecución del análisis:

- devtools
- rtracklayer
- GenomicRanges
- ChIPQC
- DT
- dplyr
- tidyR

- ChIPseeker
- TxDb.Hsapiens.UCSC.hg19.knownGene
- soGGi
- limma
- venn
- Rsubread
- csaw
- edgeR
- UpSetR
- enrichR

La totalidad el código utilizado para ambos análisis se encuentra disponible en un repositorio GitHub (<https://github.com/javi-rey/Trabajo-fin-de-master>)

3. Efecto del Ibrutinib en la expresión génica de PBMC vía scRNA-seq

Lectura y evaluación de los datos

Procedimos a la lectura de los datos generados y pretratados por los autores del artículo original utilizando Cell ranger . Se utilizó el paquete Seurat cuyo primer paso es la generación de un objeto de clase “seurat”.

Partimos de un total de 42675 células, entre todos los tipos celulares y todas las muestras. Hemos seleccionado genes que hayan sido detectados en, al menos, 10 células (16423 sobre 19105). Pasamos a la fase de control de calidad: Analizamos para cada muestra el número de genes reconocidos (nFeature_RNA), el número de *counts* (nCount_RNA) y el porcentaje de genes mitocondriales detectados (percent.mt), Figura 3.

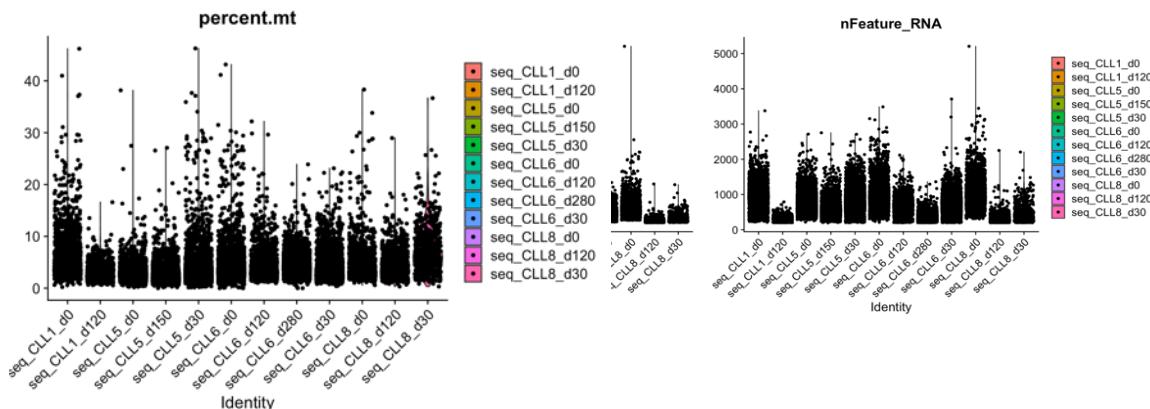


Figura 3: Gráficas de evaluación de la calidad de las muestras scRNA-seq utilizadas.

Observamos que hay tres muestras con bajos números de *counts*, aunque valores no demasiado bajo de genes detectados. Decidimos no eliminarlos dado el reducido número de muestras disponibles. En cuanto a los valores de ADN mitocondrial, todas las muestras tienen valores inferiores a un 10% en la mayoría de sus células. Decidimos comprobar la correlación entre número de *counts* y de genes, así como la correlación entre estos números y la cantidad de ADN mitocondrial detectada (Figura 4).

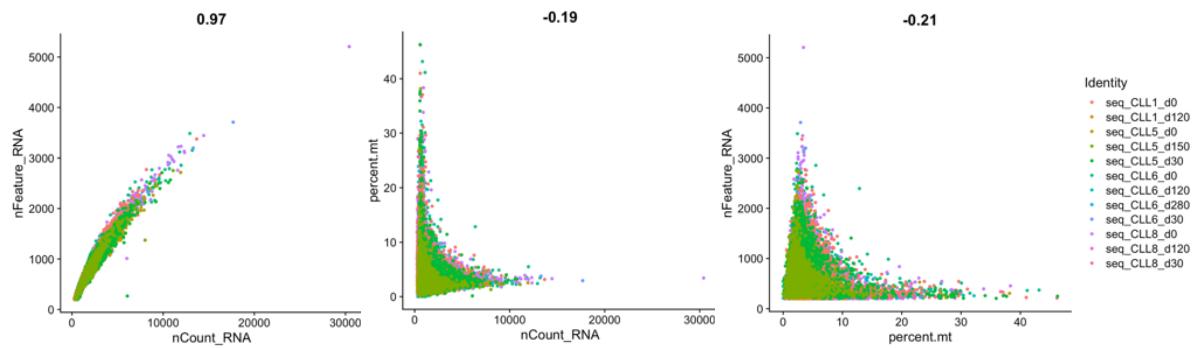


Figura 4: Correlación entre índices de calidad de las muestras. Coeficientes de correlación (r) incluidos en el título de cada gráfica.

Comprobamos la esperada alta correlación entre Counts y Features ($r=0,97$) y también que las células con mayor número de estos tienen bajos niveles de ADN mitocondrial (correlación negativa con $r=-0,19$ y $r=-0,21$), lo que indica que la mayor parte de las células analizadas están en buen estado. Decidimos aplicar un criterio similar al recomendado por los creadores de Seurat, también seguido por los autores el artículo referencia del estudio, seleccionando células con ADN mitocondrial inferior al 15%, más de 200 genes detectados y menos de 2500, lo que excluirá células en mal estado así como su ausencia o la presencia de dos células en la reacción. Este proceso nos deja con 42246 células (un 99%).

Anulación de los batch effects y selección de genes de interés

Para evitar batch effects observados en análisis anteriores realizados durante el desarrollo de las PECs, decidimos normalizar y encontrar las variables de interés de manera independiente para cada muestra. A continuación, pasamos a la fase de normalización de los counts para cada muestra, utilizando la recomendada por seurat. A continuación decidimos seleccionar los genes de interés, que son aquellos con alta variabilidad de expresión entre diferentes células. Seleccionamos 2000 genes de interés comunes a todas las muestras, destacando los 20 con mayor variabilidad detectada (Figura 5). Tras ello realizamos la integración de los datos de todas las muestras.

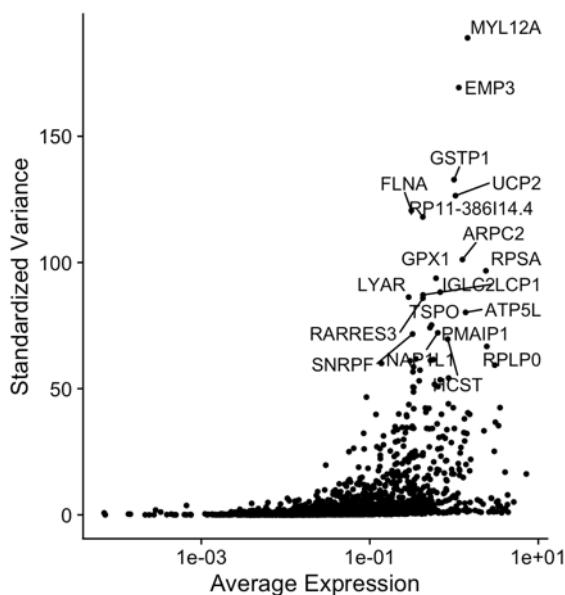


Figura 5. Genes de interés. Proyección de los 2000 genes seleccionados en función de su expresión y variabilidad. Destacados los 20 con mayor variabilidad

A continuación escalamos los valores de los genes seleccionados con el fin de que los genes altamente expresados no tuvieran un mayor peso en los procesos de reducción de dimensionalidad y clustering posteriores.

Análisis de componentes principales (PCA) y elección de dimensiones

Tras ello pasamos a realizar un análisis inicial de componentes principales. A continuación se muestran los diez genes con más peso (positivo o negativo) en cada una de las 5 primeras componentes principales.

```

## PC_1
## Positive: NKG7, IL32, CCL5, CD3D, ITGB2, GZMA, S100A4, ANXA1, CTSW, GZMH
## Negative: IGKC, IGHM, TCL1A, HLA-DRA, CD74, CD79A, MALAT1, MZB1, ADTRP, CD79B

## PC_2
## Positive: CST3, FCN1, LYZ, CSTA, AIF1, S100A9, RP11-1143G9.4, SERPINA1, LST1, FCER1G
## Negative: CCL5, IL32, GZMH, GZMA, CTSW, CD3D, NKG7, TRGC1, CST7, CD3E

## PC_3
## Positive: MALAT1, MT-CO1, CTB-50L17.16, CEP70, PIEZO1, S100A8, NEAT1, GABPB1-AS1, KLF2, VCAN
## Negative: RPS2, RPL10, RPS18, RPL15, RPL3, RPS19, RPS3, RPS12, RPLP1, RPLP0

## PC_4
## Positive: CDKN1C, FCGR3A, LINC01272, MS4A7, RRAS, HES4, CKB, TCF7L2, BATF3, HCK
## Negative: VCAN, CD14, S100A8, RP11-1143G9.4, S100A12, CSF3R, MS4A6A, ALDH2, S100A9, FOS

## PC_5

```

```
## Positive: TRDC, GZMB, GNLY, FGFBP2, PRF1, GZMH, KLRD1, TRGC1, CLIC3, NKG7
## Negative: IL7R, TNFRSF25, GZMK, LTB, CD40LG, CD28, RP11-664D1.1, AQP3, GPR183,
RCAN3
```

A continuación observamos la distribución de las células en las primeras dos dimensiones, comprobando que ninguna muestra se alejaba globalmente del resto y que todas contenían *clusters* celulares similares (Figura 6). En el gráfico observamos los genes con mayor peso en cada una de las dimensiones utilizadas.

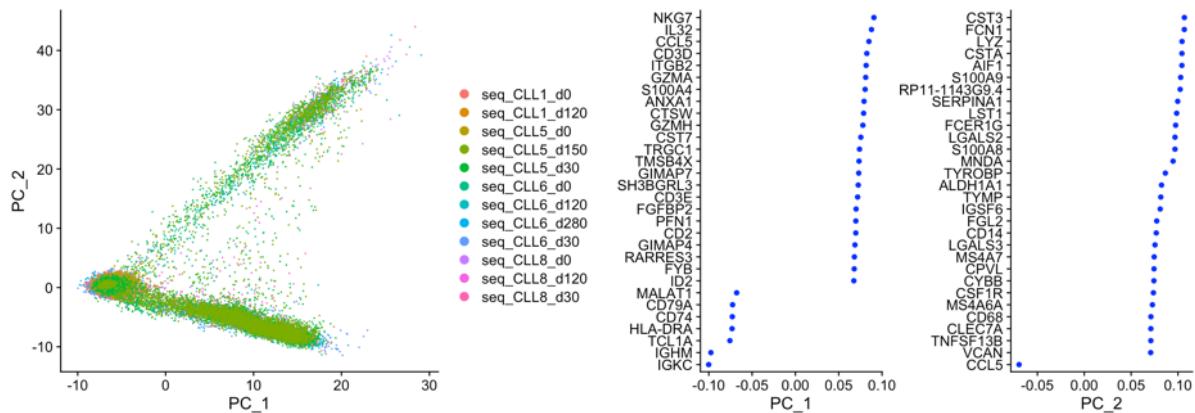


Figura 6: Distribución de las células de cada muestras en las primeras dos componentes y genes con mayor peso en la distribución.

Observamos también el reparto de la expresión de los principales genes de las primeras 10 componentes (Figura 7), lo que nos sirvió más tarde para decidir cuantas utilizamos en los siguientes procesos

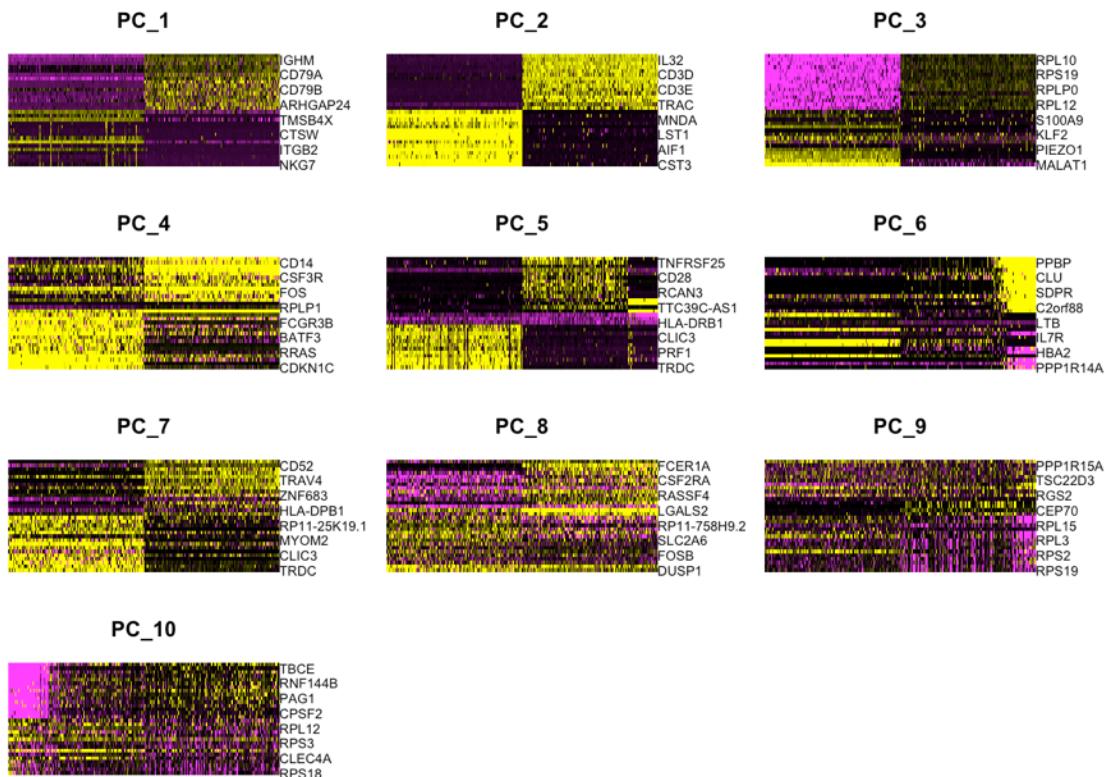


Figura 7: Expresión de los principales genes de las primeras 10 componentes. Se seleccionaron 500 células al azar para cada gráfica.

A continuación utilizamos dos gráficos para seleccionar cuantas dimensiones seleccionamos: un Elbow plot y un estudio JackStraw.

En el primer gráfico (Figura 8)

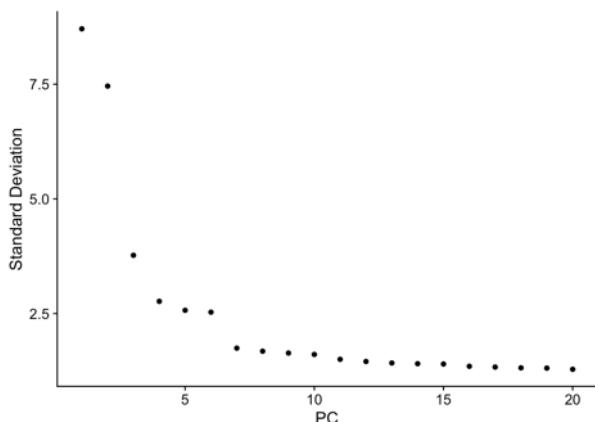


Figura 8. Elbow plot

El gráfico nos muestra que, al menos, las primeras 10 a 15 dimensiones pueden ser relevantes.

Como alternativa realizamos un cálculo jackStraw según el cual las primeras 20 dimensiones serían significativas, decidimos reducir el número a 10 ya que éste número parece justificado en ambos casos y nuestro objetivo no necesitaba identificar un alto número de subpoblaciones que se diferenciasen en dimensiones ulteriores.

Clustering de poblaciones

Pasamos a la detección de las diferentes poblaciones celulares que componen las muestras, para lo cual utilizamos el método Louvain, implementado en Seurat. El objetivo es comprobar si nuestro método consigue distinguir entre los diferentes tipos celulares que componen las muestras de sangre periférica. Tras varios ensayos, utilizamos *FindClusters(combined, resolution = 0.15)*. La baja resolución elegida nos evitó la generación de un exceso de *clusters/poblaciones*. Se detectaron 7 poblaciones diferentes.

Representamos los grupos hallados en las dos primeras dimensiones del PCA (Figura 9)

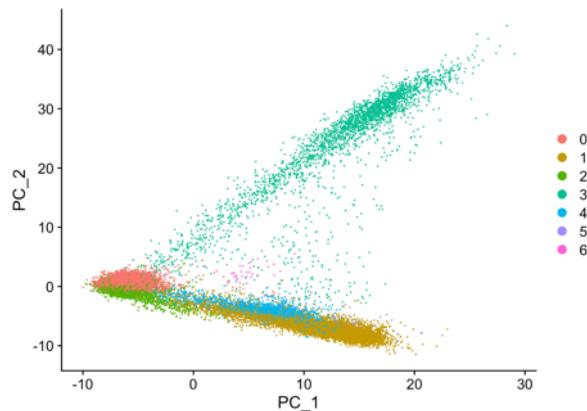


Figura 9. Distribución de las poblaciones detectadas; representación por PCA.

Las poblaciones detectadas parecían solapadas en la representación, por ello realizamos un estudio UMAP , utilizando las primeras 10 componentes(Figura 10). También comprobamos la distribución de las muestras dentro de la distribución UMAP.

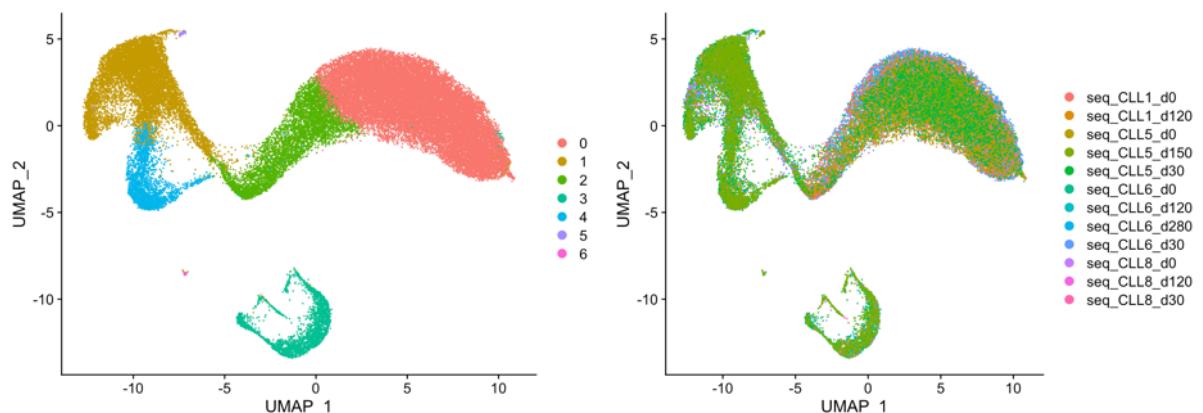


Figura 10. Distribución de las poblaciones detectadas; representación por UMAP.

El método UMAP nos permite separar mejor los *clusters* de manera gráfica. Observamos un número de poblaciones conveniente así como un solapamiento perfecto de las muestras entre sí, lo que demuestra lo apropiado del tratamiento contra los *batch effects*. Sin la corrección de *batch effects* obteníamos *clusters* comunes mientras que existían *clusters* asociados a todas las muestras de un mismo paciente.

División de las muestras en antes y durante el tratamiento e identificación de las poblaciones celulares

Decidimos incorporar la información relativa al tratamiento con Ibrutinib. El bajo número de muestras, así como la falta de coincidencia en la temporalidad de las mismas, hizo que decidieramos dividir las muestras en Ibrutinib SI (8) e Ibrutinib NO (4).

A continuación decidimos comprobar lo apropiado del clustering e identificar los tipos celulares a los cuales correspondía cada cluster. Procedimos a buscar los principales marcadores de cada población detectada utilizando *FindAllMarkers*. Como ejemplo, se muestran los primeros 5 marcadores de cada cluster, ordenados en función de su valor p-ajustado (Tabla III).

Principales marcadores de los clusters detectados						
0	1	2	3	4	5	6
IGKC	GNLY	MALAT1	S100A8	GZMK	PCNA	PPBP
IGHM	NKG7	MT-ND3	S100A9	IL7R	STMN1	PF4
TCL1A	CCL5	MT-ND2	LYZ	IL32	HMGB2	GNG11
CD79A	GZMH	ARHGAP24	CST3	TRBC1	TUBB	CLU
HLA-DRA	GZMA	GABPB1-AS1	FCN1	GIMAP7	TUBA1B	NCOA4

Tabla III. Principales marcadores de los *clusters* detectados.

Utilizamos los marcadores mostrados en la tabla anterior, así como otros clásicos para definir las poblaciones que conforman las muestras. Para ello nos servimos de dos gráficos que ilustran la expresión de los marcadores en los diferentes clusters. En el primero (Figura 11), también incorporamos la información relativa a si la muestra proviene de un paciente en tratamiento o no.

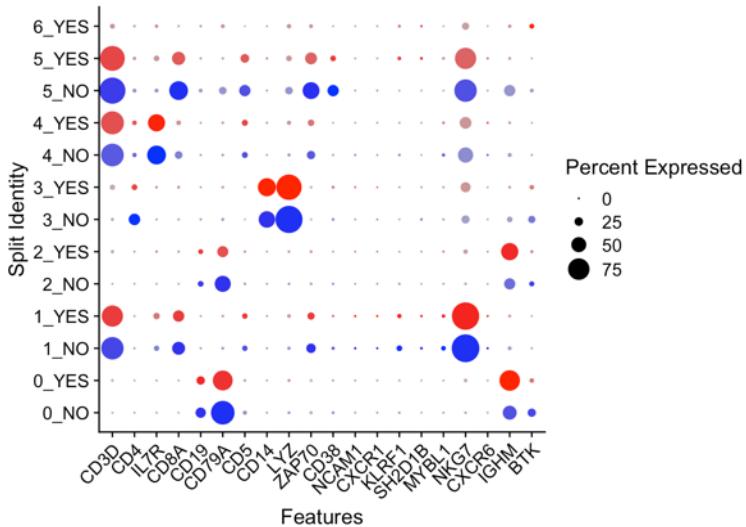


Figura 11: Expresión de marcadores en los diferentes clusters y efecto del Ibrutinib

El segundo gráfico muestra la expresión relativa de los marcadores en el gráfico UMAP (Figura 12).

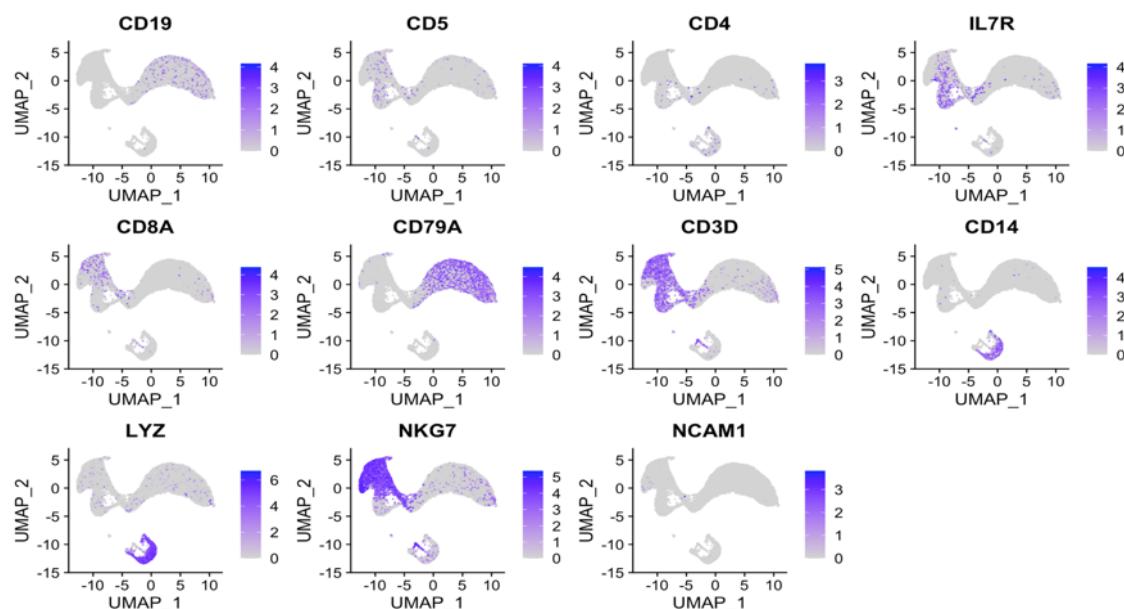


Figura 12: Expresión relativa de marcadores; UMAP.

La combinación de estos gráficos y el estudio de la función de los genes seleccionados nos permitió asignar un tipo celular a los diferentes clusters, excepto para el cluster 5, minoritario, de difícil asignación y poco interesante para nuestro estudio

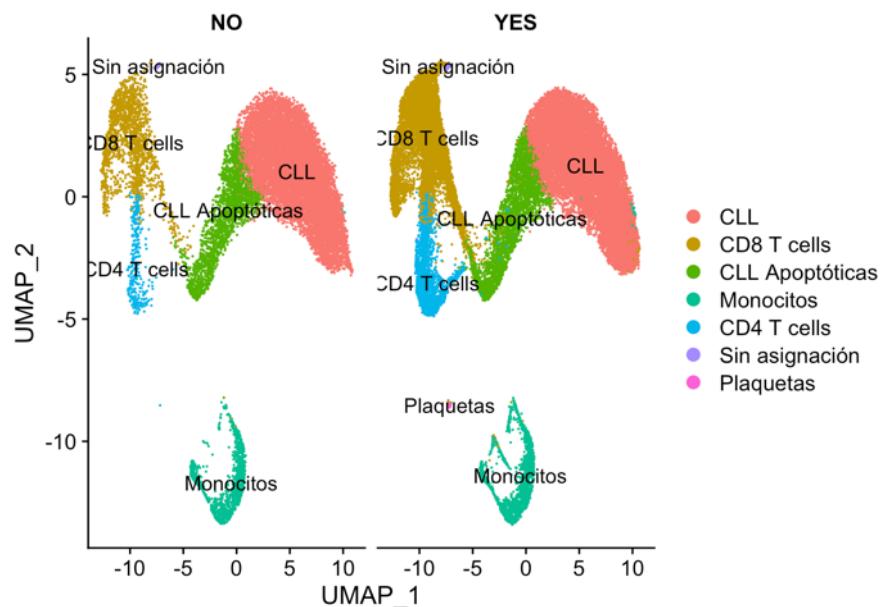


Figura 13: Asignación de poblaciones a los *clusters* detectados; UMAP.

La separación de muestras en función el tratamiento nos permite comenzar a explorar sus efectos. Podemos comprobar que, globalmente, el predominio de células tumorales en ausencia de Ibrutinib se ve algo reducido ante el tratamiento, con una mayor proporción de células T. Decidimos explorar la distribución de poblaciones muestra a muestra (Figura 14). Podemos comprobar las diferencias en las cantidades y proporciones celulares entre pacientes. Además, podemos constatar el cambio, de proporciones por efecto del Ibrutinib, más acusado en unos pacientes que en otros.

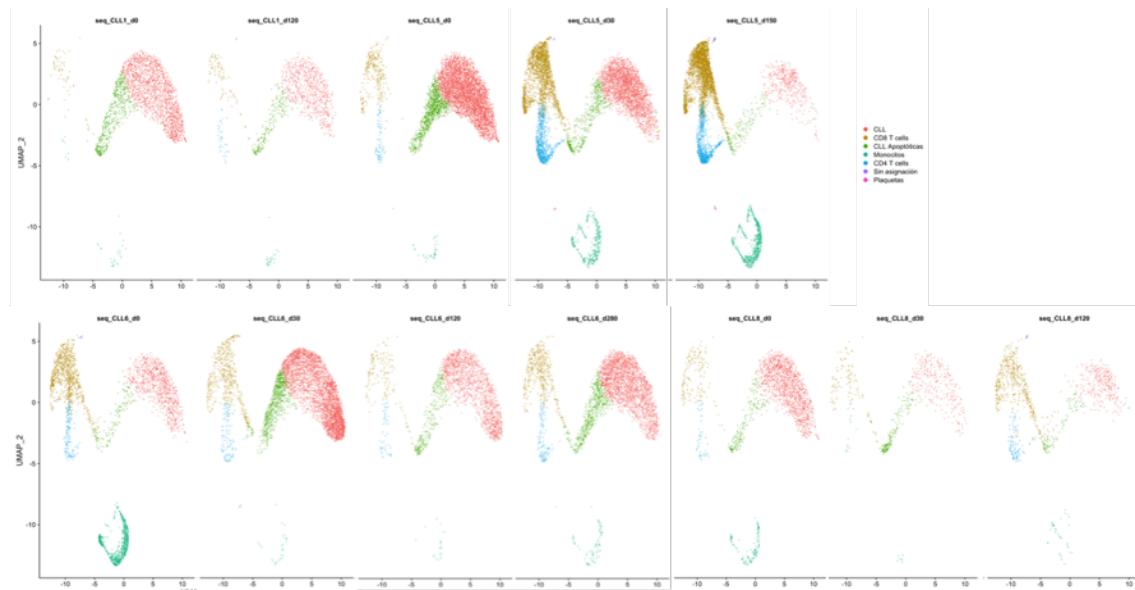


Figura 14: Asignación de poblaciones, UMAP muestra a muestra.

Análisis de expresión diferencial por ibrutinib y Gene Ontology

Células tumorales

Comenzamos seleccionando las células tumorales y encontrando los principales genes con expresión diferencial de las condiciones con y sin tratamiento gracias a *FindMarkers*.

Encontramos más de 200 genes inhibidos por el tratamiento, así como más de 50 genes sobreexpresados. Los resultados fueron representados usando un HeatMap (Figura 15) y almacenados en tablas junto con la información estadística asociada (tabla anexa).

Células tumorales (CLL)

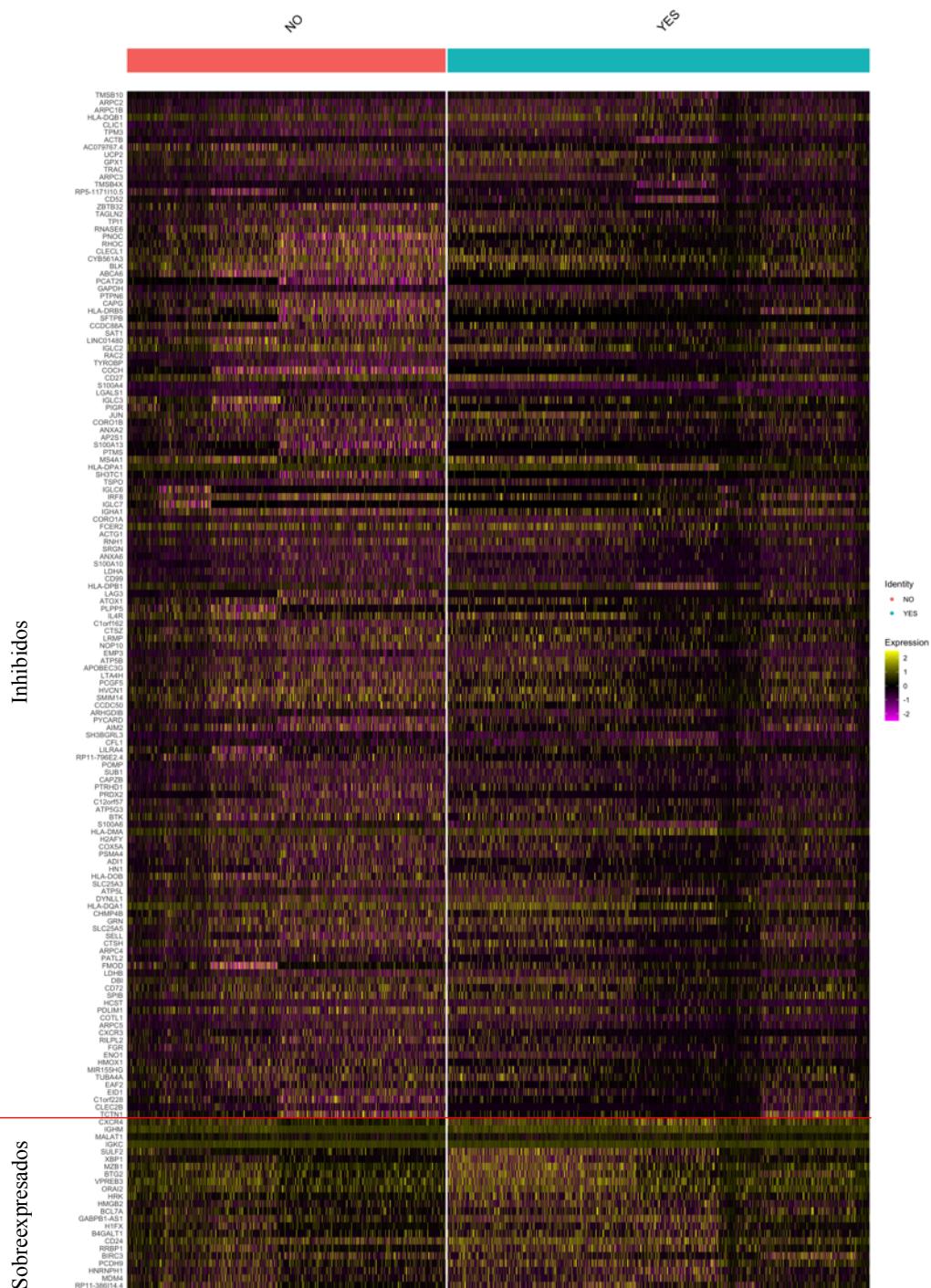


Figura 15: Genes con expresión diferencial por efecto del tratamiento con ibrutinib en células tumorales.

A continuación realizamos un análisis de *Gene Ontology* (GO) utilizando la aplicación web EnrichR utilizando la base de datos Panther 2016. El objetivo fue comprobar si existía una alteración general de la expresión de genes asociados a adhesión y migración celulares.

En primer lugar introdujimos la lista de genes sobreexpresados, sin embargo ninguno de los términos GO destacados tenía un valor p ajustado relativo al enriquecimiento adecuado (tabla anexa).

Tras ello, testamos la lista de genes inhibidos obteniendo varios términos fuertemente enriquecidos, entre los cuales destacaron *Cytoskeletal regulation by Rho GTPase Homo sapiens* y *Integrin signalling pathway Homo sapiens*, además de dos términos asociados a la inhibición de la activación de células B vía BCR (Figura 16 y tabla anexa).

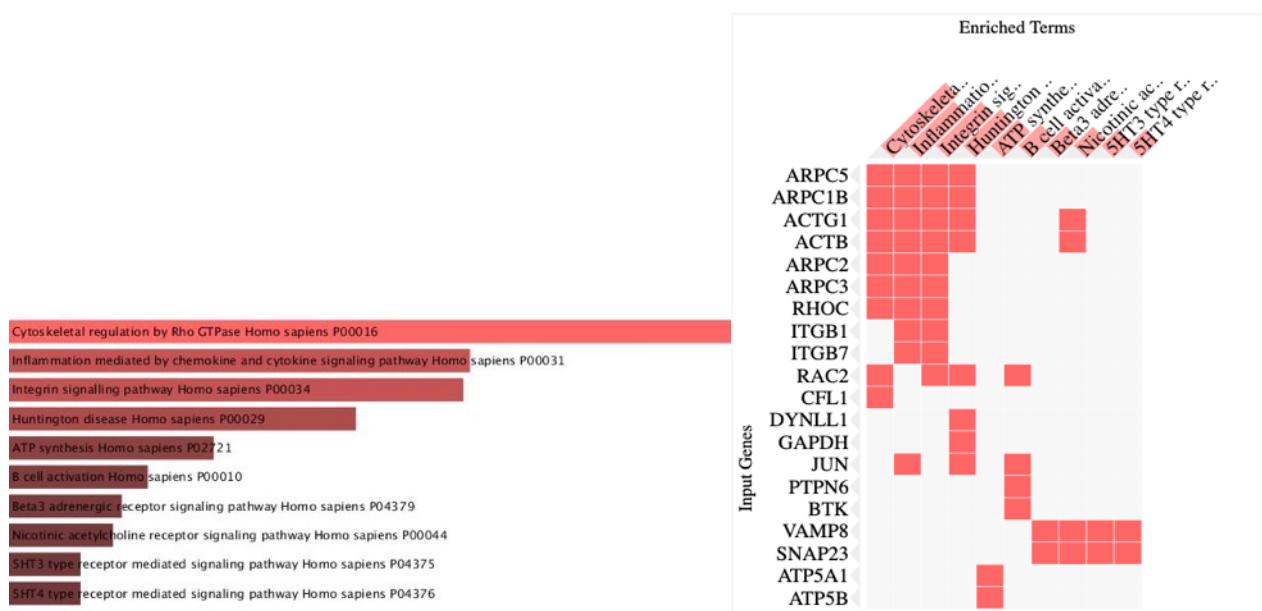


Figura 16. Análisis GO con EnrichR en células tumorales. La tabla de la izquierda muestra los términos GO destacados, ordenados en función su menor valor p-ajustado. La tabla de la derecha muestra los genes responsables de la selección de los términos.

Decidimos comprobar uno a uno los niveles de expresión de algunos de los genes de interés para nuestro estudio (Figura 17)

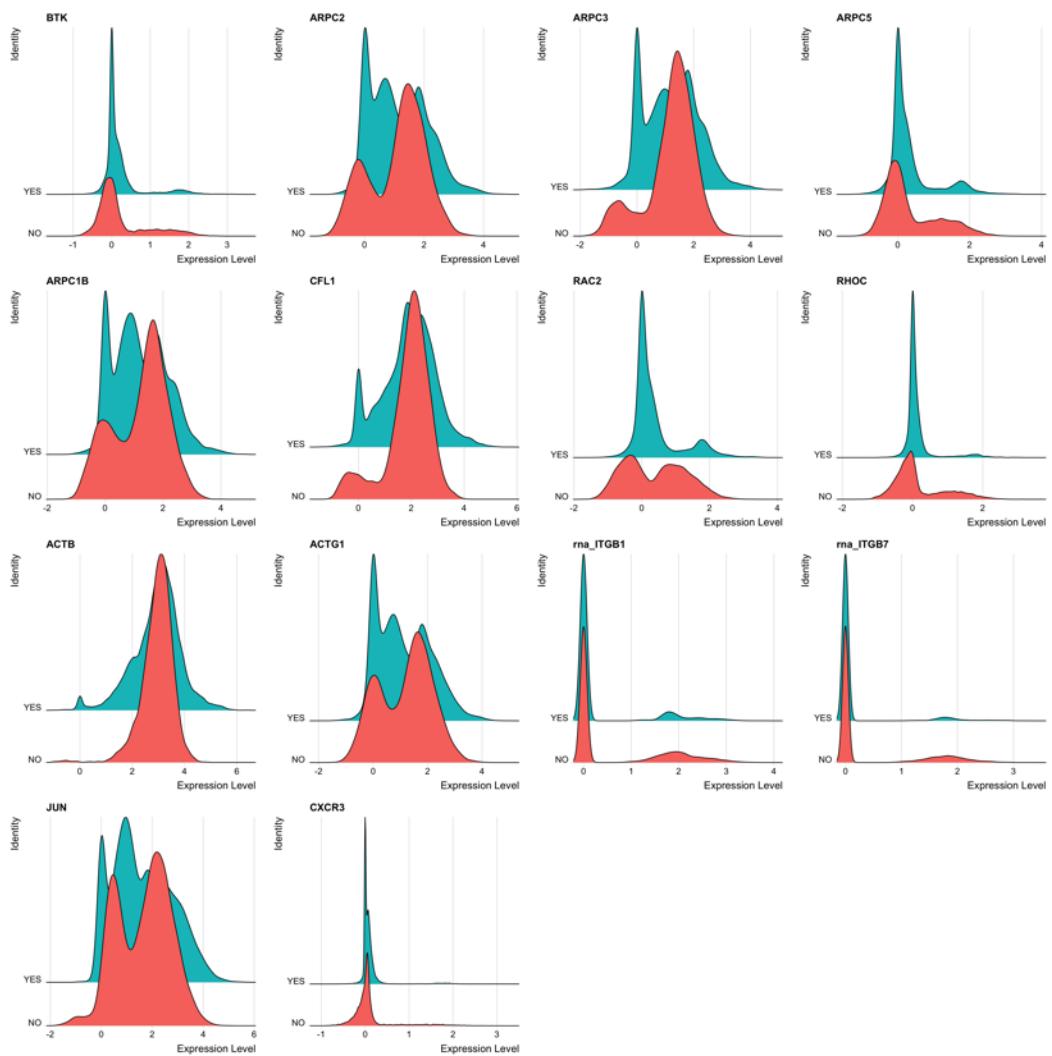


Figura 17. Efecto del Ibrutinib en la expresión relativa de genes de adhesión y migración de células tumorales.

De estos resultados podemos concluir que existe una inhibición generalizada de la maquinaria asociada a la adhesión y migración celular, con un impacto en los genes del complejo ARP 2/3, la inhibición de las integrinas beta 1, responsable de la adhesión, entre otros, a colágeno y fibronectina, y beta 7, responsable de la adhesión a laminina. Por otra parte, se observa también efecto sobre varias RhoGTPasas, las cuales se encargan e coordinar procesos como la polimerización y contracción del citoesqueleto. Estas proteínas suelen actuar en respuesta a señales provenientes de receptores de chimioquinas e integrinas, con lo que los resultados parecen coherentes entre sí. En el apartado de genes sobreexpresados, destaca CXCR4, asociado a la migración a médula ósea.

Linfocitos T y monocitos

A pesar de que pudimos detectar otras poblaciones celulares y de que los autores del estudio original analizaron el efecto del Ibrutinib en células NK y linfocitos B no tumorales, decidimos restringir la evaluación de los efectos colaterales a linfocitos T CD4 y CD8 así como monocitos (CD14). Existen tres razones para tomar esta decisión:

- La dificultad de encontrar una resolución de *clustering* que nos identificara estas poblaciones sin dividir en exceso las ya identificadas.
- La complejificación del proyecto debido al aumento de poblaciones a analizar
- El bajo interés del estudio de estas poblaciones para integrar los resultados a aquellos obtenidos en el estudio COMPUTREAT-CLL.

Para el estudio de las 3 poblaciones inmunes se utilizó el mismo método que el usado anteriormente para las células LLC. A continuación se muestran los principales resultados:

Linfocitos T-CD8

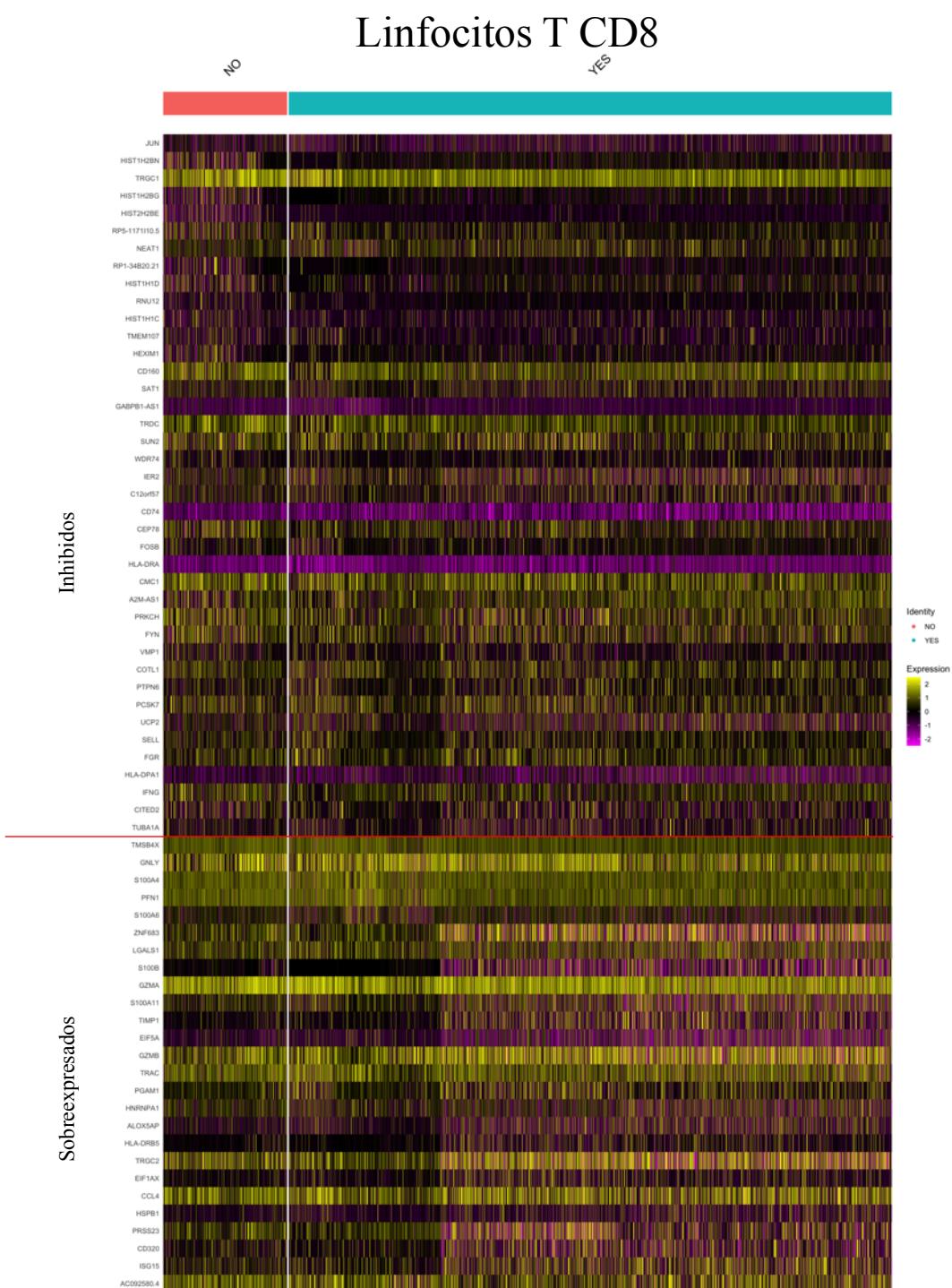
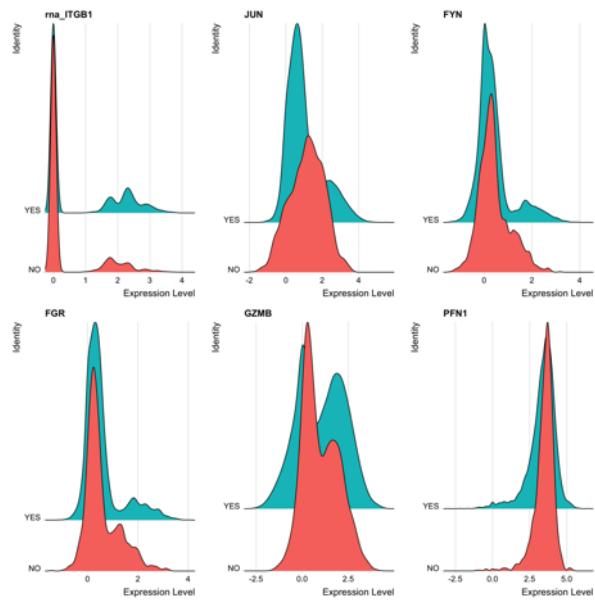


Figura 18: Genes con expresión diferencial por efecto del tratamiento con ibrutinib en linfocitos T CD8

A diferencia de las células LLC, los linfocitos T CD8 no mostraron una inhibición global de los genes de adhesión y migración (Figura 18). Como ya fue descrito por Rendeiro y colaboradores, los principales genes alterados tras tratamiento se relacionaron con la activación (JUN, inhibido) y función lítica de estas células (como los genes GZMB o PFN1, que codifican componentes de los gránulos líticos, sobreexpresados). Sorprendentemente observamos un aumento de la expresión de ITGB1. También destaca la inhibición de las kinasas de la familia Src: FYN y FGR, las cuales participan en la reorganización del citoesqueleto ante señales externas (quimioquinas y adhesión, pero también reconocimiento de antígeno). Es por ello que una inhibición de la capacidad de migración de linfocitos T no puede ser totalmente excluida. A modo de ejemplo, se incluyen gráficos de expresión de los genes



mencionados (Figura 19).

Figura 19: Efecto del Ibrutinib en la expresión relativa de genes de interés en linfocitos T CD8.

Linfocitos T CD4

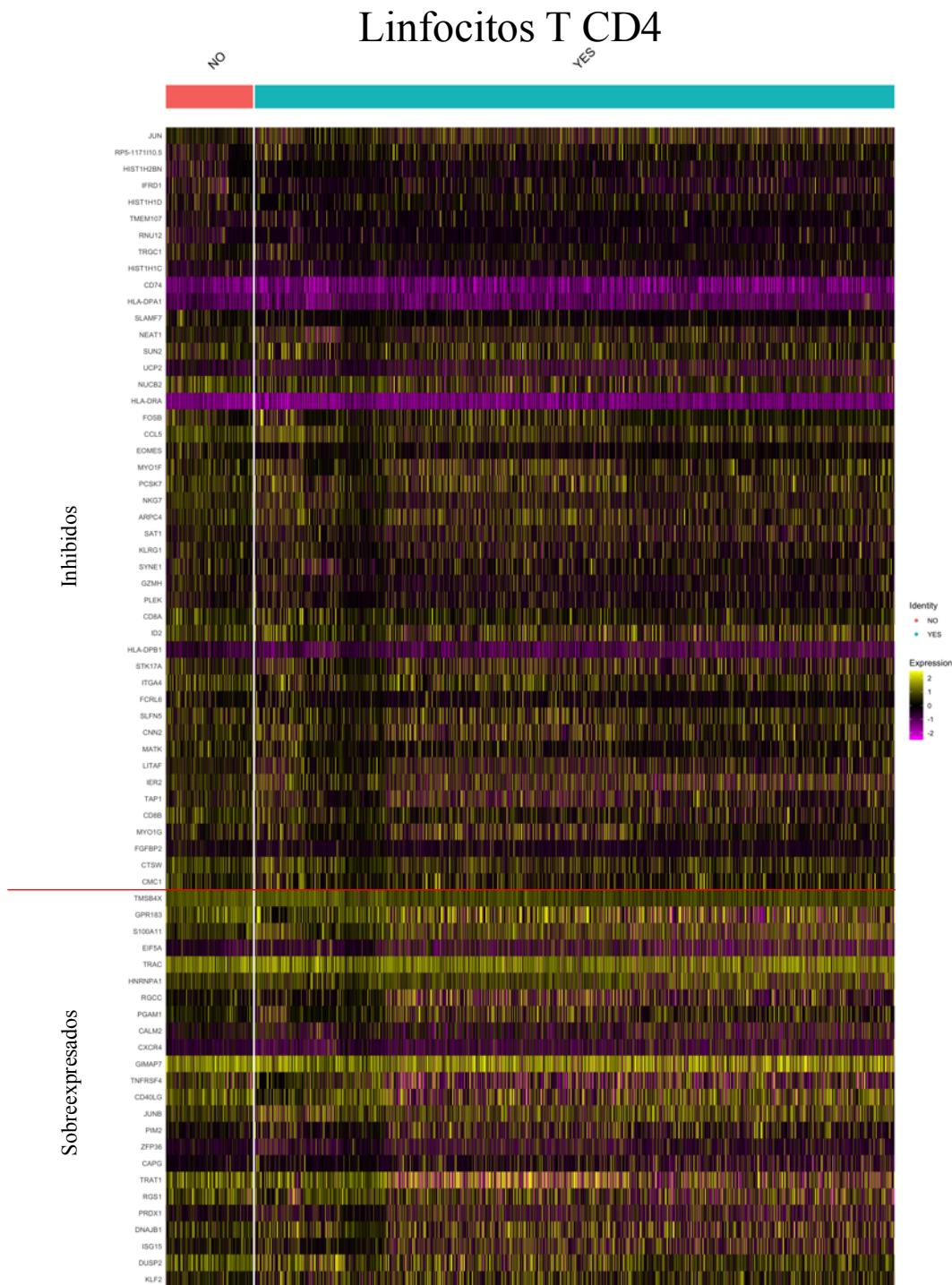


Figura 20 Genes con expresión diferencial por efecto del tratamiento con ibrutinib en linfocitos T CD4.

El efecto del Ibrutinib en linfocitos T CD4 tuvo un patrón mucho menos marcado (Figura 20) y no especialmente asociado a genes de citoesqueleto o integrinas. Destaca la inhibición de JUN, responsable de la activación de estas células, compartida con los T CD8 y también encontrada por los autores del estudio original.

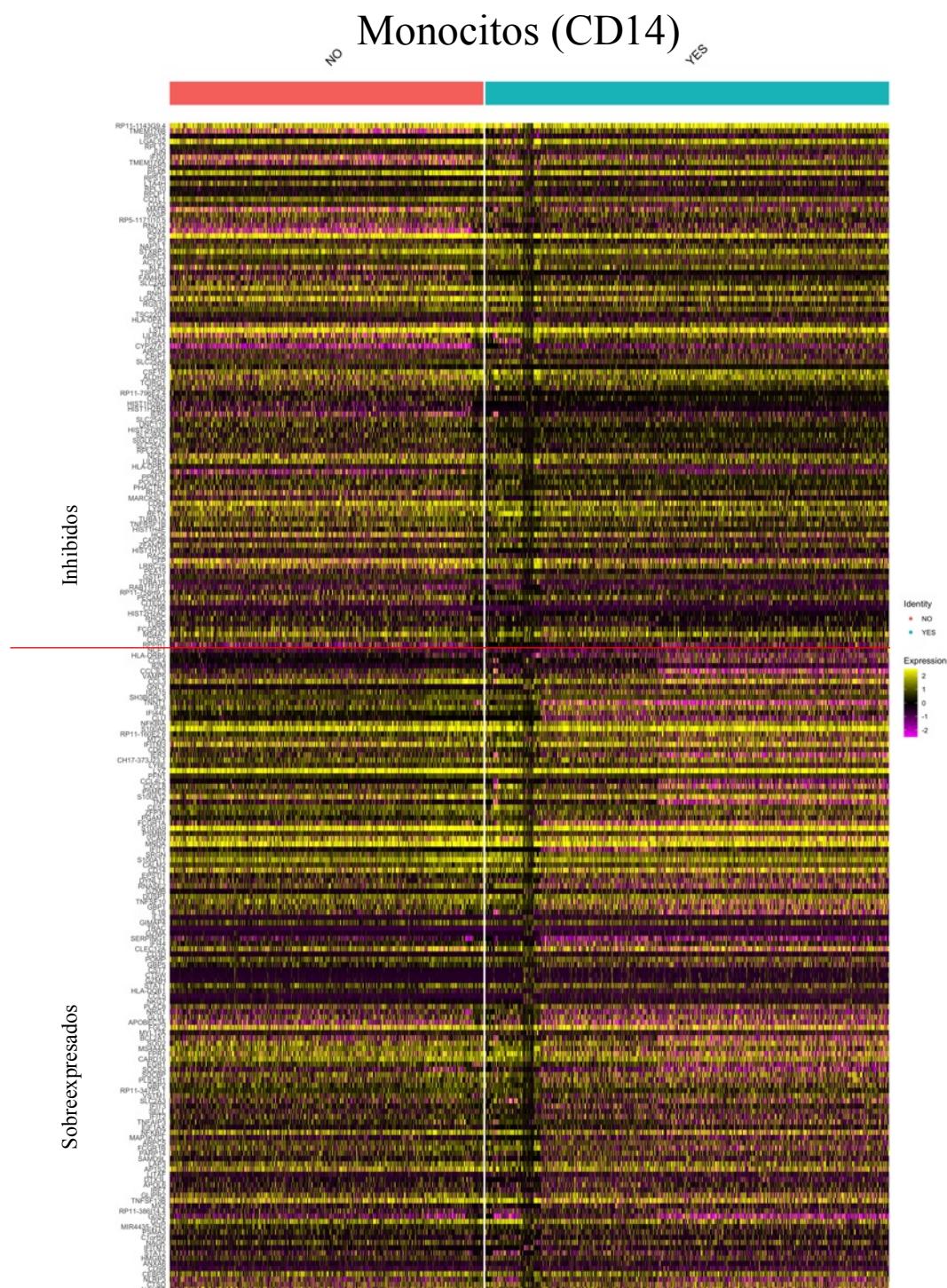
Monocitos (CD14)

Figura 21 Genes con expresión diferencial por efecto del tratamiento con ibrutinib en monocitos.

En monocitos, el análisis GO mostró un aumento de la respuesta asociada a NFkB de estas células ante tratamiento con Ibrutinib (Figura 21). Esto se manifiesta en el enriquecimiento detectado de los términos *Interferon-gamma signaling pathway*, *Inflammation mediated by chemokine and cytokine signaling pathway Homo sapiens* y *Toll receptor signaling pathway*.

En cuanto a los genes inhibidos, los términos asociados a citoesqueleto volvieron a aparecer, así como genes asociados al complejo ARP 2/3 y RhoGTPasas (Figura 22).

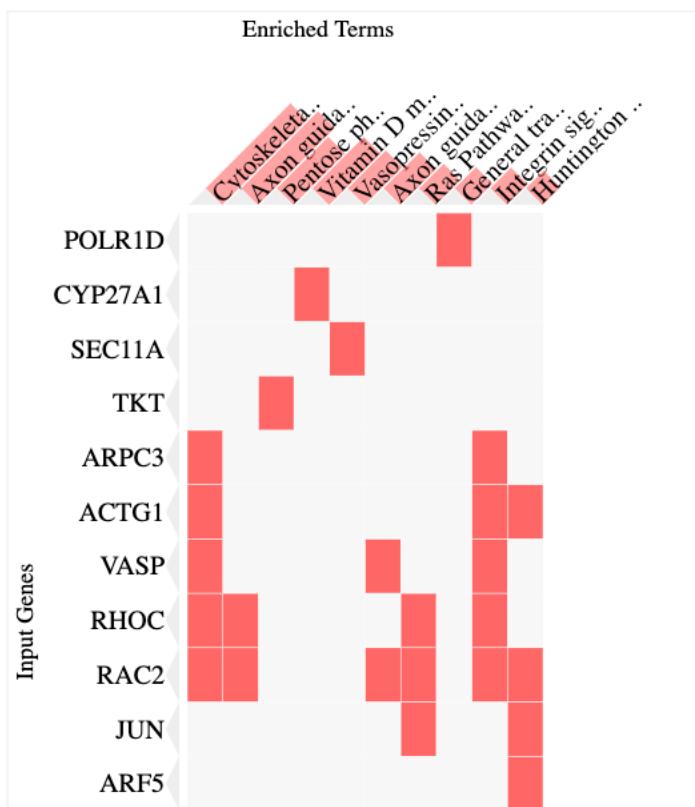


Figura 22: Principales términos GO y genes asociados, inhibidos por Ibrutinib en monocitos.

Globalmente, el análisis de datos scRNA nos ha permitido comprobar que el abandono de las células LLC de los tejidos linfoides secundarios en respuesta a Ibrutinib puede explicarse, al menos en parte, por un silenciamiento generalizado de la expresión de genes asociados a los proceso de adhesión y migración.

Por otra parte, hemos podido comprobar que estos efectos no son comunes a todos los tipos celulares, pero quizás compartidos por los monocitos. Los efectos colaterales del Ibrutinib han quedado plasmados en una reducción e la capacidad de activación de células T, así como en el cambio global del microentorno tumoral hacia un perfil más proinflamatorio.

4. Efecto del Ibrutinib en la apertura de la cromatina de células immunes vía ATAC-seq.

El análisis fue similar en todos los tipos celulares, aunque realizado de forma individual. El estudio utilizó las muestras procedentes de células tumorales (CLL), linfocitos T CD4, linfocitos T CD8 y Monocitos. Las razones para ello fueron similares a las que restringieron el análisis del scRNA-seq. A continuación se muestra el proceso detallado tomando como ejemplo las muestras de células tumorales.

Efecto en células tumorales (CLL)

Lectura y carga datos

En primer lugar, se estudió el formato de los archivos compartidos por Rendeiro y colaboradores. Se observó que para cada muestra existían dos archivos: Un primer archivo .BigWig con la información de lecturas y su localización en la cromatina, y un segundo archivo .NarrowPeak, generado tras el proceso de *Peak Calling* realizado por los autores con MACS2. Seleccionamos este último tipo de archivos con el fin de simplificar el procesado de datos. Utilizamos el paquete *rtracklayer*. Dado que las muestras procedían de células aisladas por citometría, cada tipo celular poseía sus propios archivos.

Filtrado de picos detectados

Filtramos los datos para mantener sólo aquellos pertenecientes a los cromosomas canónicos y no sexuales. Eliminamos también los picos artefactales asociados a “blacklists”. Para ello se excluyeron los picos asociados a regiones contenidas en la lista “The DAC Blacklisted Regions”, del proyecto ENCODE.

Anotación de picos

A continuación pasamos a identificar, en cada muestra, la asociación de los picos detectados a los diferentes genes del genoma humano, así como su rol más probable. Utilizamos la herramienta *ChIPseeker* y la base de datos de anotación *TxDb.Hsapiens.UCSC.hg19.knownGene*. Tomamos como límite de asignación a un gen una distancia de 3000 bases antes o tras el inicio de la transcripción (Tss). Decidimos incluir un gráfico que mostrara, para cada muestra, la proporción de elementos reconocidos. Esto nos ayudó a detectar muestras aberrantes (Figura 23)

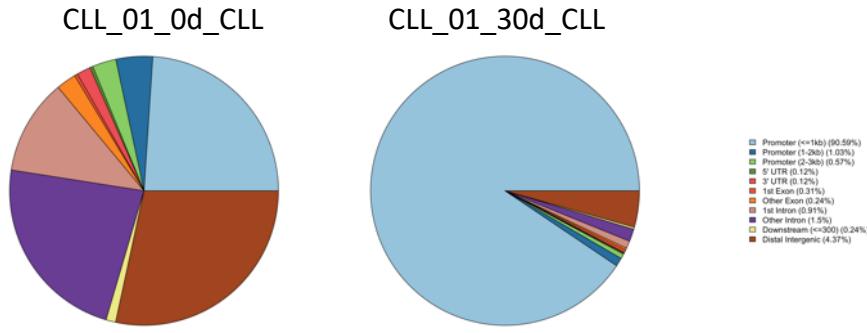


Figura 23: Resumen de la anotación de picos. A la izquierda, una muestra equilibrada, a la derecha, una muestra aberrante

Observamos que el piechart de la muestra CLL1_30d_CLL era muy distinto al resto, lo que nos hizo sospechar que los datos de la muestra podrían ser erróneos.

Estudio diferencial del efecto del ibrutinib. Búsqueda de picos comunes.

Analizamos cuál era el grado de solapamiento de los picos entre muestras, y realizamos una primera observación del agrupamiento de picos en muestras del mismo tipo. Establecimos 5 categorías en función de los días de tratamientos: Día 0, Día 3 (entre 1 y 3), Día 8, Día 30 y Día 120 (120 o más días de tratamiento)

En un inicio realizamos el análisis usando diagramas de Venn. Esto nos permitió detectar la diferencia entre CLL1_30d_y el resto de muestras de su tiempo(Figura 24). Finalmente, la falta de proporcionalidad de éstos, así como el elevado número de muestras por categoría hizo que nos decantáramos por gráficos Upset (Figura 25).

Estos volvieron a mostrar la

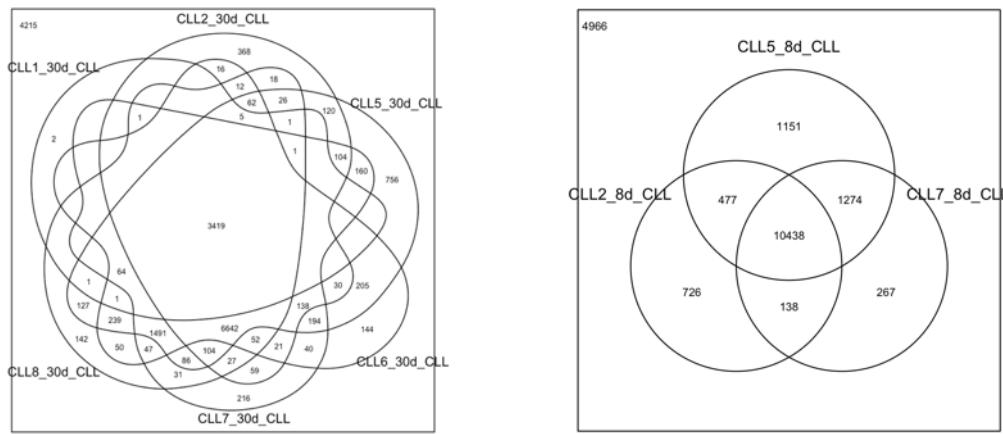


Figura 24. Coincidencia de picos entre muestras de CLL. Diagramas de Venn.

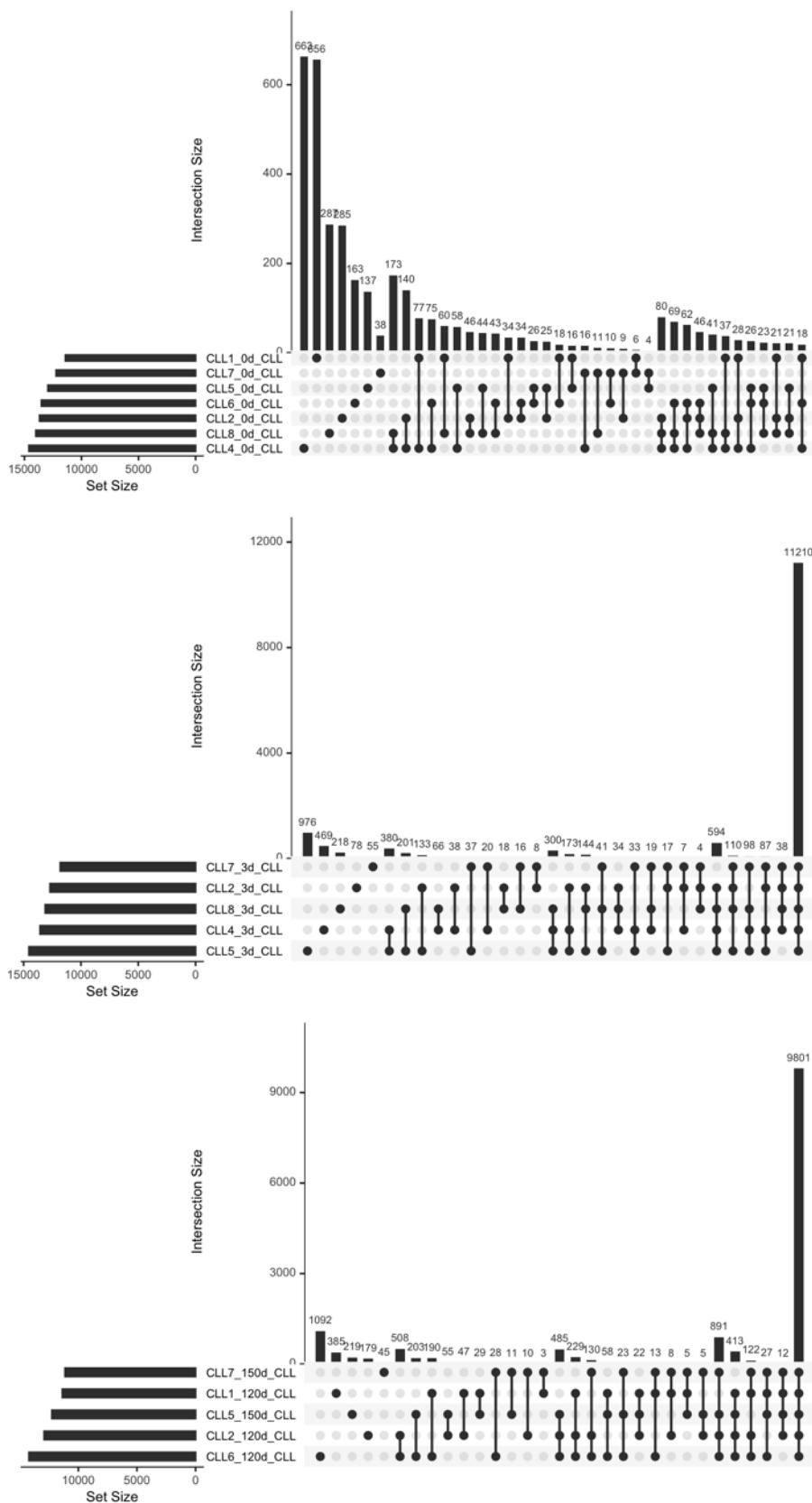


Figura 25. Coincidencia de picos entre muestras de CLL. Gráficos Upset.

En el resto de grupos de muestras, no parece haber ninguna que destaque sobre el resto.

Análisis de componentes principales

Realizamos una reducción de dimensiones por PCA (Figura 26), observando que la muestra CLL1_30d vuelve a ser diferente al resto. Por ello, decidimos eliminar esta muestra y volver a realizar el cálculo de picos consenso y la reducción de dimensiones. Tras ello, agrupamos las muestras en 5 categorías temporales y observamos la distribución de las mismas en función del tipo de paciente. Pudimos observar la agrupación de muestras de un mismo paciente, lo que denota la presencia de *batch effects* o, más probablemente, una alta variabilidad entre pacientes.

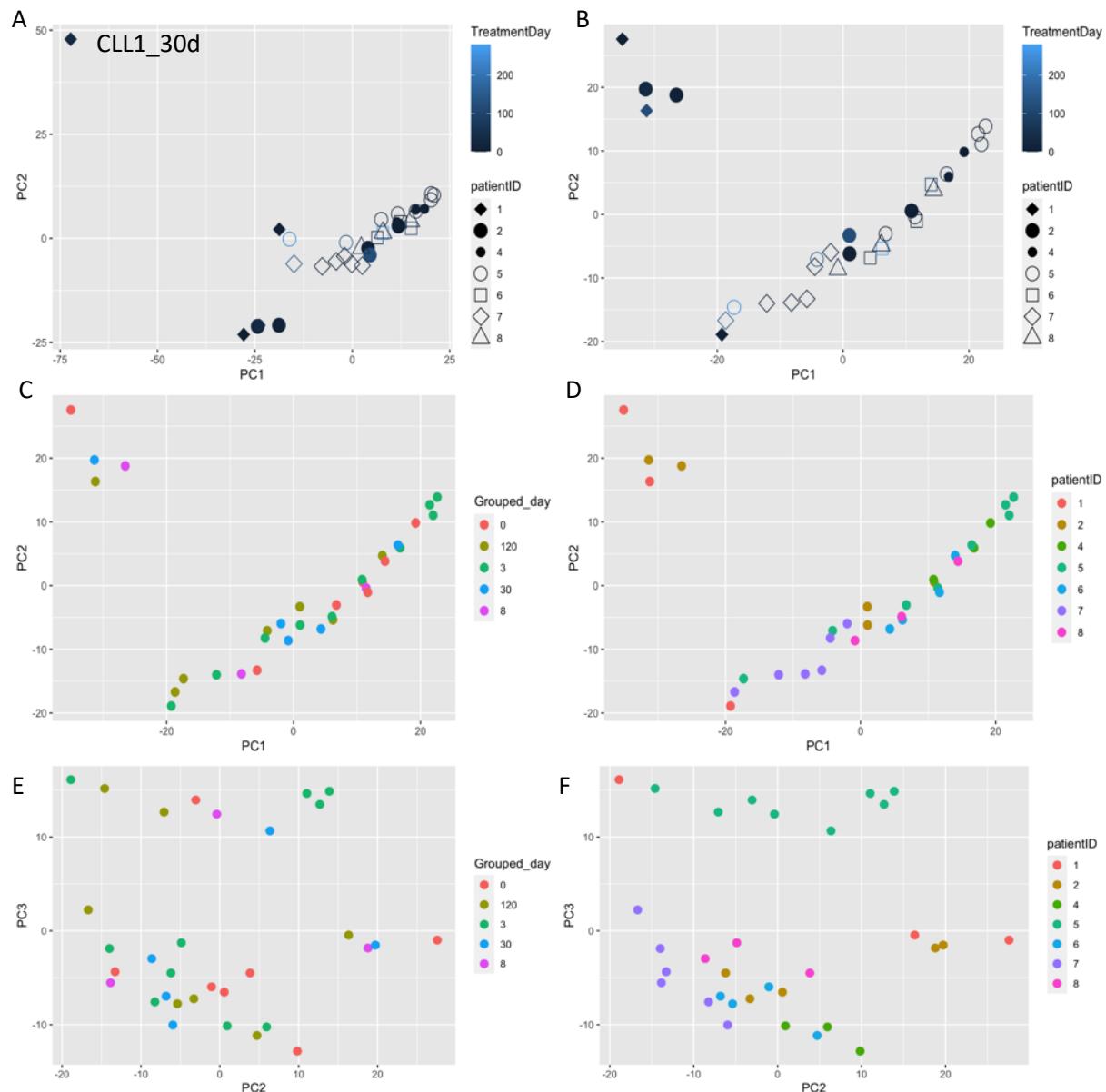


Figura 26: Componentes principales. A) Distribución del total de muestras. B) Distribución tras la eliminación de CLL1_30d. C) Distribución tras agrupación de muestras. D) Distribución con respecto al paciente. E y F) Distribuciones en PC2 y PC3.

Los autores del estudio original decidieron sustraer la primer componente como tratamiento de corrección de *batch effects*. Decidimos comprobar qué distribución compondrían las muestras en la combinación de PC2 y PC3. Pudimos observar que los

batch effects están presentes también en la PC3, por lo que el método no nos pareció suficientemente apropiado.

Definición del grupo de picos consenso asociados a promotores

Evaluamos la cantidad de picos comunes entre muestras, con el fin de establecer un conjunto de los mismos compartidos entre varias de éstas, evitando "outliers". Decidimos conservar los picos compartidos en, al menos, 5 muestras (13996 picos), número que representa el número de pacientes medio por grupo de muestras. Tras ello, acotamos la distancia absoluta máxima de los picos al TSS a 1000 bases con el objetivo de conservar sólo los picos asociados a promotores génicos.

Establecimiento de un índice de apertura de la cromatina en promotores con respecto al tiempo de tratamiento.

Decidimos pasar al análisis diferencial de la apertura de la cromatina en función del tratamiento con Ibrutinib. Para ello utilizamos las muestras agrupadas en 5 categorías temporales. Tras ello consideramos que la presencia de un pico asociado al promotor del gen indicaba la apertura del mismo en esta muestra, asignándole un valor de 1 para ese gen. Si el pico no estaba en la muestra, consideramos que el promotor se encontraba "cerrado" (con un valor de 0). Para cada grupo de muestras, establecimos un índice de 0 a 1 para cada pico consenso calculado como su "apertura media" para el mismo. Véase como ejemplo el siguiente caso:

PEAK NUMBER	SYMBOL	GENENAME	0	3	8	30	120
2280	STIM1	stromal interaction molecule 1	1	0.9	0.66	0.4	0.14

Aquí, todas las muestras antes del tratamiento contienen el pico, luego consideramos que el promotor el gen STIM1 se encontraba abierto. El tratamiento fue progresivamente bajando el número de muestras que contenían el pico, hasta acabar un índice de 1/7 en el grupo de 120 días o más.

Detección de promotores afectados por el tratamiento con ibrutinib y evolución temporal

Utilizando este índice, pasamos a la selección de promotores inhibidos por efecto el Ibrutinib así como aquellos activados por el mismo. Para ello seleccionamos los genes con promotores abiertos a día 0 (con un índice superior a 0,7) y, dentro de estos, seleccionamos aquellos que mostraban un valor inferior a 0,3 en alguno de los otros grupos. Realizamos también la operación inversa (0,3 o menos a día 0 y 0,7 o más en alguno de los otros grupos) a fin de detectar promotores activados por efecto del fármaco.

Los resultados fueron recogidos en forma de tablas anexas y *heatmaps* (Figura 27)

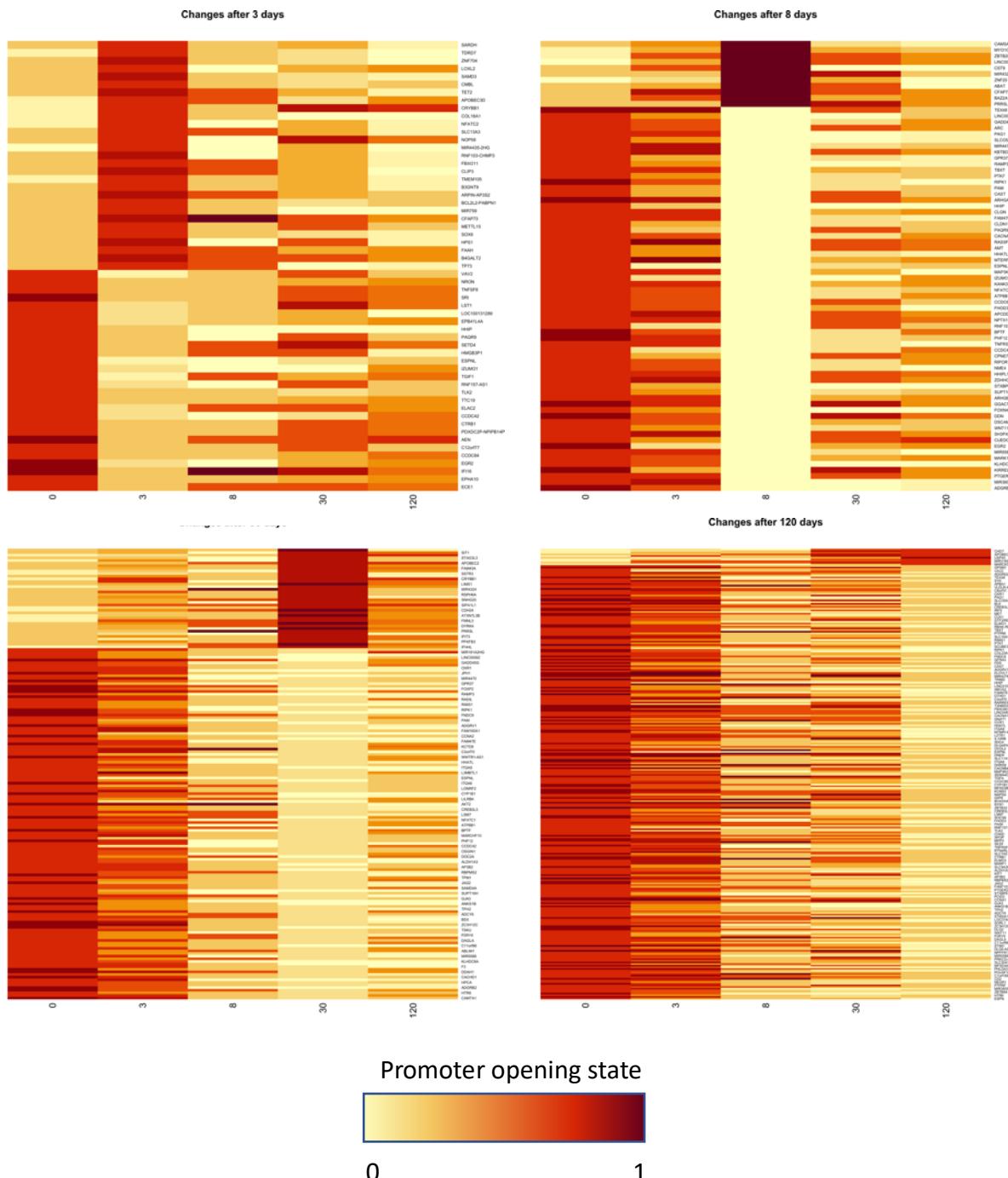


Figura 27: Cambios en la apertura de promotores de células LLC por efecto del tratamiento con Ibrutinib

Una primera observación de los genes destacados nos muestra fuertes diferencias con los resultados obtenidos en el scRNAseq, con una ausencia de gran parte de genes asociados a la señalización dependiente de integrinas o asociada a quimioquinas. Pasamos a realizar un análisis de *Gene Ontology* para observar los procesos afectados específicamente por el tratamiento.

Estudio GO de los genes con promotores diferencialmente accesibles

Con el objetivo de comprobar cuáles son las principales funciones celulares cuyos promotores se ven alterados por el tratamiento con Ibrutinib, Decidimos realizar un estudio de *Gene Ontology* utilizando el paquete *EnrichR*, comprobando las bases de datos *GO_Molecular_Function_2018*, *GO_Biological_Process_2018* y *Panther_2016*.

Recuperamos para todas las situaciones (silenciamiento o apertura de los promotores tras 3, 8, 30 o 120 días de tratamiento) los 10 primeros términos destacados. Con respecto a la base de datos Panther 2016, utilizadas anteriormente para evaluar los resultados de scRNA-seq. Decidimos representar s términos GO enriquecidos en el caso de promotores silenciados por el tratamiento (Tabla XXX).

Term	Overlap	P,value	Adjusted,P,value	Genes
Open at day 0, closed at day 3				
B cell activation Homo sapiens P00010	1/57	0,0768538	1	VAV2
T cell activation Homo sapiens P00053	1/73	0,0973821	1	VAV2
Endothelin signaling pathway Homo sapiens P00019	1/75	0,0999169	1	ECE1
PDGF signaling pathway Homo sapiens P00047	1/112	0,1455923	1	VAV2
Open at day 0, closed at day 8				
B cell activation Homo sapiens P00010	2/57	0,0143678	1	MAP3K2;NFATC1
De novo pyrimidine ribonucleotides biosynthesis Homo sapiens P02740	1/9	0,0284395	1	NME4
Vasopressin synthesis Homo sapiens P04395	1/10	0,0315499	1	PAM
De novo pyrimidine deoxyribonucleotide biosynthesis Homo sapiens P02739	1/10	0,0315499	0,8833977	NME4
Salvage pyrimidine ribonucleotides Homo sapiens P02775	1/10	0,0315499	0,7067181	NME4
De novo purine biosynthesis Homo sapiens P02738	1/26	0,0800034	1	NME4
Beta2 adrenergic receptor signaling pathway Homo sapiens P04378	1/28	0,0858897	1	CACNA1D
Beta1 adrenergic receptor signaling pathway Homo sapiens P04377	1/28	0,0858897	1	CACNA1D
Oxytocin receptor mediated signaling pathway Homo sapiens P04391	1/39	0,1176079	1	CACNA1D
SHT2 type receptor mediated signaling pathway Homo sapiens P04374	1/46	0,1372256	1	CACNA1D
Open at day 0, closed at day 30				
Angiogenesis Homo sapiens P00005	7/142	4,85756E+01	0,005440472	JAG2;DLL4;NOTCH1;AKT2;F2R;PRKCA;F3
Notch signaling pathway Homo sapiens P00045	3/38	2,08998E+03	0,117038762	DLL4;JAG2;NOTCH1
Apoptosis signaling pathway Homo sapiens P00006	4/102	4,93418E+03	0,184209406	AKT2;PRKCA;RIPK1;ATF3
Endothelin signaling pathway Homo sapiens P00019	3/75	1,39828E+04	0,391517689	AKT2;PRKCA;ADCY6
Blood coagulation Homo sapiens P00011	2/38	2,67639E+04	0,5995122	F2R;F3
Oxytocin receptor mediated signaling pathway Homo sapiens P04391	2/39	2,80889E+04	0,524325175	CACNA1D;PRKCA
p53 pathway feedback loops 2 Homo sapiens P04398	2/45	3,65678E+04	0,585084734	CCNA2;AKT2
SHT2 type receptor mediated signaling pathway Homo sapiens P04374	2/46	3,80656E+04	0,532917797	CACNA1D;PRKCA
2-arachidonoylglycerol biosynthesis Homo sapiens P05726	1/6	3,95373E+04	0,492019513	DAGLA
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway Homo sapiens P00026	3/127	5,38526E+04	0,603148824	HTR6;CREB3L3;ADCY6
Open at day 0, closed at day 120				
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway Homo sapiens P00026	7/127	0,001501897	0,1682124	GPSM1;HTR6;CREB3L3;PRKAR1B;CREB3L2;ARRB1;ADCY6
Beta2 adrenergic receptor signaling pathway Homo sapiens P04378	3/28	0,00583688	0,3268653	RYR1;CACNB4;CACNA1D
Beta1 adrenergic receptor signaling pathway Homo sapiens P04377	3/28	0,00583688	0,2179102	RYR1;CACNB4;CACNA1D
B cell activation Homo sapiens P00010	4/57	0,006777025	0,1897567	MAP3K2;SYK;NFATC1;VAV2
Inflammation mediated by chemokine and cytokine signaling pathway Homo sapiens P00031	7/188	0,01262127	0,2827164	CCR1;NFAT5;AKT2;NFATC1;ARRB1;SOC6;ADCY6
Transcription regulation by bZIP transcription factor Homo sapiens P00055	3/42	0,017872604	0,3336219	CREB3L3;PRKAR1B;CREB3L2
p53 pathway feedback loops 2 Homo sapiens P04398	3/45	0,021465781	0,3434525	CCNA2;CCNA1;AKT2
SHT2 type receptor mediated signaling pathway Homo sapiens P04374	3/46	0,02274476	0,3184266	CACNB4;CACNA1D;KCNK3
Angiogenesis Homo sapiens P00005	5/142	0,040327563	0,5018541	JAG2;NOTCH1;AKT2;F2R;F3
T cell activation Homo sapiens P00053	3/73	0,072165292	0,8082513	AKT2;NFATC1;VAV2

Tabla IV: Términos GO destacados por PANTHER en la inhibición de promotores por efecto del Ibrutinib.

Pudimos observar que tras 3 u 8 días de tratamiento, todavía no existe un patrón claro de genes cuyo promotor se vea afectado, a diferencia de los resultados tras 1 o 4 meses. Esto coincide con la interpretación de los resultados hecha por los autores del estudio, la cual habla de un programa de quiescencia que se instala progresivamente en estas

células. Tras un mes, observamos múltiples vías de señalización afectadas como las asociadas a Notch1, receptor de tipo 5HT2 y proteínas g heterotriméricas. Estos 2 últimos términos, así como el gen NOTCH1, también se encuentran entre los 10 más destacados tras 4 meses, junto a la señalización vía receptores beta adrenérgicos, la recepción de citoquinas y, sobre todo, la activación de células B, diana principal del tratamiento con Ibrutinib.

Estos resultados coinciden los descritos por Rendeiro y colaboradores y validan en cierto modo nuestro análisis, sin embargo distan de los obtenidos en el análisis por scRNA-seq. El único término GO asociado al citoesqueleto (*Inflammation mediated by chemokine and cytokine singaling pathway*) aparece tras 120 días de tratamiento, mucho después de la salida masiva de células de los órganos linfoides.

Junto a éste destacamos la presencia de Vav2 (activador de Rho-GTPasas) y AKT2, central en la recepción de quimioquinas.

Efecto en linfocitos T y en monocitos

Decidimos aplicar a las células CD4, CD8 y a los monocitos el mismo pipeline ejecutado previamente para células tumorales. No nos vemos obligados a eliminar ninguna muestra y obtenemos picos afectados por el tratamiento con Ibrutinib. Sin embargo, la observación de los genes de forma individual, así como el estudio GO no hayan un especial efecto del tratamiento en la apertura de promotores de genes asociados al citoesqueleto o recepción de quimioquinas. Esto, unido al menor interés e estas poblaciones para nuestro estudio y lo reducido del tiempo restante para la entrega de la memoria, hace que los resultados sólo sean presentados en forma de tablas anexas.

5. Conclusiones

El presente estudio ha intentado comprender el modo de acción del Ibrutinib a través de la explotación de los resultados de una serie de experimentos scRNA-seq y ATAC-seq realizado en pacientes de LLC tanto antes, como a diferentes puntos de tratamiento. La hipótesis principal del mismo establecía que la salida masiva de células tumorales se debía a una alteración de los genes de adhesión y migración de las mismas por efecto del fármaco.

Los resultados derivados del análisis transcriptómico de 12 muestras nos permiten afirmar que el tratamiento con Ibrutinib genera un silenciamiento de diferentes genes del complejo ARP 2/3, responsables de los cambios en la forma celular necesarios para la migración celular(Wu et al., 2012), la inhibición de las integrinas beta 1, y beta 7, responsables de la adhesión a los principales componentes de la matriz extracelular. (Huttenlocher & Horwitz, 2011)

La inhibición de estos dos procesos ocurre también en su control por parte de las RhoGTPases, ya que varios de sus miembros también se encuentran inhibidos, lo que refuerza la consistencia de los resultados obtenidos.

En el apartado de genes sobreexpresados, destaca CXCR4, asociado a la migración a médula ósea. Este resultado, en contra de nuestra hipótesis, contrasta con el descenso en la capacidad de migración hacia CXCL12, así como la disminución de la presencia en membrana celular encontrada en COMPUTREAT-CLL y otros estudios (Chen et al., 2016).

Esto nos alerta de la limitación en la interpretación de resultados basados sólo en cantidades de RNA mensajero y apunta a que los genes de interés anteriormente mencionados deberían ser validados individualmente en futuros ensayos.

En segundo lugar, quisimos evaluar los efectos colaterales del Ibrutinib en otras células inmunes, centrándonos en los procesos de migración de éstas. Esto fue debido a la observación de una fuerte reducción de la capacidad quimiotáctica de células T por efecto del Ibrutinib (datos preliminares de COMPUTREAT-CLL). En este caso, no obtuvimos un perfil de expresión alterado específicamente para adhesión o migración en ninguna de las poblaciones de células T. Los principales efectos observados fueron una disminución general e la activación de CD8 (JUN) y la adopción de un perfil efector citotóxico (genes de gránulos líticos). Los únicos genes de interés inhibidos por Ibrutinib fueron dos quinasas de la familia Src, sin embargo su implicación en otros procesos de activación aparte de la migración celular los vuelve candidatos poco atractivos. Por su parte, las células CD4 no parecieron afectadas en exceso por el tratamiento.

A diferencia de los linfocitos T, los monocitos si mostraron una inhibición de ciertos genes asociados a ARP2/3 y de RhoGTPasas. Esto, unido a un aumento de la expresión de genes diana de NFkB podría indicar un cambio en la predisposición a la diferenciación de los monocitos hacia un perfil de tipo M1, una célula de carácter proinflamatorio y con baja capacidad de migración.

La promoción de este tipo de macrófagos, unida al aumento del carácter lítico de los linfocitos T CD8 podría ayudar a destruir los nichos de supervivencia de las células tumorales.

Clave en la validación de nuestro protocolo de análisis, nuestros resultados coincidieron en gran medida con los mostrados por los autores del estudio original, a pesar de las diferencias entre protocolos, como en la fase de corrección de *batch effects*.

Es importante resaltar también los objetivos que no pudieron cumplirse en este apartado. En primer lugar, uno de los objetivos contemplaba estudiar la heterogeneidad de las células LLC antes y durante el tratamiento, con el fin de establecer si existía una selección de ciertas subpoblaciones y si éstas tenían un perfil de migración característico. Unos primeros pasos se dieron en el estudio, aumentando la resolución del clustering, sin embargo, los primeros resultados fueron poco esperanzadores. Esto y la necesidad de avanzar el análisis ATAC-seq hizo que decidieramos abandonar esta parte del estudio.

En segundo lugar, en el plan de trabajo preveíamos separar los pacientes en al menos dos clases en función del aumento de la linfocitosis en respuesta a Ibrutinib. Según nuestra hipótesis, podrían existir perfiles diferenciales de expresión de genes asociados a migración que explicaran estas diferencias. Sin embargo, el bajo número de pacientes y muestras (4 y 12, respectivamente), unido a la variabilidad entre pacientes, hizo poco viable un estudio de este tipo.

En el apartado referente al estudio del efecto del Ibrutinib en la apertura de la cromatina mediante ATAC-seq, decidimos retomar los datos procesados por los autores tras el *Peak Calling*. Tuvimos problemas en la lectura inicial de los archivos, debido al poco conocimiento de los formatos y de *Genomic Ranges*. Además, no pudimos encontrar un valor cuantitativo fiable de la intensidad de señal de los picos. Esto retrasó el proyecto y nos obligó a elegir un sistema semi-cuantitativo basado en la presencia-ausencia de los picos en cada muestra.

Los resultados obtenidos utilizando este método mostraron que a partir del mes de tratamiento se produce un silenciamiento general de múltiples vías de señalización asociadas a supervivencia y activación e células leucémicas. Esto coincide plenamente con lo descrito por los autores originales, que hablan de la “adopción progresiva de un estado de quiescencia”. Resulta de especial interés el silenciamiento del promotor de NOTCH1 por efecto del tratamiento, ya que este gen suele estar mutado y constitutivamente activo en CLL, con lo que el Ibrutinib revertiría los efectos de estas mutaciones (Rosati et al., 2018).

Como hemos comentado anteriormente, la aparición de algunos genes aislados, así como de un término GO, asociados a adhesión y migración, se produce días más tarde del éxodo masivo de células de los tejidos linfoides, con lo cual no pueden explicar este fenómeno.

Esta falta de concordancia entre los datos procedentes de ambas técnicas impidió la integración e ambos resultados. Que los resultados de ATAC-seq no expliquen aquellos obtenidos mediante scRNA-seq, aun cuando ambos análisis parecen correctos puede deberse a razones técnicas o biológicas:

- La elección el análisis de promotores dejó fuera la posibilidad de otras regulaciones, como la ejercida por elementos *enhancers*.
- La regulación de la transcripción puede realizarse sin afectar a la apertura de la cromatina mediante el reclutamiento diferencial e factores de transcripción.
- El reconocimiento de poblaciones fue diferente en ambos métodos: en el scRNA-seq fue decisión del investigador basándose en la interpretación de la

expresión de ciertos marcadores. En el caso de ATAC-seq, las células fueron separadas previamente en base a marcadores de citometría estandarizados, más fiables que el método anterior. Por ello, no pueden excluirse cambios entre ambas asignaciones.

- Los datos procedían de un bajo número de réplicas, especialmente en el caso del scRNA-seq (12). En el ATAC-seq, se analizó una media de 5 muestras por punto temporal.

Al margen de estas explicaciones, existía en ambos juegos de datos un fuerte “efecto paciente”. Este efecto fue corregido en el protocolo de scRNA-seq mediante el tratamiento independiente de cada muestra y su posterior integración.

El efecto también es claro en las muestras de ATAC-seq y ya fue descrito por Rendeiro y colaboradores. Sin embargo, al testar su método de corrección basado en la eliminación de la primera componente observamos que las componentes 2 y 3 también contenían estos efectos. Tras discutirlo con mi tutora, decidimos intentar una aproximación diferente mediante la utilización de un modelo lineal con la fórmula *Open~PatientID+Ibrutinib*, para después extraer la variabilidad ligada sólo a Ibrutinib. Pudimos testar el modelo en el protocolo de T CD8, pero los resultados fueron similares y la falta de tiempo nos impidió profundizar más en esta vía.

Quedan, por tanto vías de mejora en ambos análisis. Por otra parte, los resultados invitan a comprobar algunas de las dianas en experimentos posteriores y sería muy interesante generar nuevas muestras de scRNA-seq de una cohorte mayor de pacientes para establecer una dinámica de los cambios transcriptómicos. También sería interesante comprobar si alguno de los pacientes presentó resistencia al tratamiento y, en tal caso, comprobar si existen perfiles de expresión distintivos (y si estos están ligados a la migración celular).

6. Glosario

LLC: Leucemia Linfocítica Crónica

CLL: *Chronic Lymphocytic Leukaemia*

Ibrutinib: Fármaco inhibidor de la quinasa Btk utilizado en clínica.

T-CD8: Subpoblación de linfocitos citotóxicos

T-CD4: Subpoblación de linfocitos “*helpers*”

CD14: Proteína de membrana que identifica a monocitos

scRNA-seq: *single cell RNA sequencing*

ATAC-seq: *Assay for Transposase-Accessible Chromatin – sequencing*

Batch effects: Variabilidad de los ensayos debida al origen de cada muestra así como a su tratamiento.

7. Bibliografía

- Allaire, J. J. (2015). RStudio: Integrated development environment for R. *The Journal of Wildlife Management*. <https://doi.org/10.1002/jwmg.232>
- Arruga, F., Gyau, B. B., Iannello, A., Deaglio, S., Vitale, N., & Vaisitti, T. (2020). Immune response dysfunction in chronic lymphocytic leukemia: Dissecting molecular mechanisms and microenvironmental conditions. In *International Journal of Molecular Sciences*. <https://doi.org/10.3390/ijms21051825>
- Burger, J. A., Ghia, P., Rosenwald, A., & Caligaris-Cappio, F. (2009). The microenvironment in mature B-cell malignancies: A target for new treatment strategies. In *Blood*. <https://doi.org/10.1182/blood-2009-06-225326>
- Chang, B. Y., Francesco, M., De Rooij, M. F. M., Magadala, P., Steggerda, S. M., Huang, M. M., Kuil, A., Herman, S. E. M., Chang, S., Pals, S. T., Wilson, W., Wiestner, A., Spaargaren, M., Buggy, J. J., & Elias, L. (2013). Egress of CD19+CD5+ cells into peripheral blood following treatment with the Bruton tyrosine kinase inhibitor ibrutinib in mantle cell lymphoma patients. *Blood*. <https://doi.org/10.1182/blood-2013-02-482125>
- Chen, S. S., Chang, B. Y., Chang, S., Tong, T., Ham, S., Sherry, B., Burger, J. A., Rai, K. R., & Chiorazzi, N. (2016). BTK inhibition results in impaired CXCR4 chemokine receptor surface expression, signaling and function in chronic lymphocytic leukemia. *Leukemia*. <https://doi.org/10.1038/leu.2015.316>
- Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., & Chang, H. Y. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics*. <https://doi.org/10.1038/ng.3646>
- Dubovsky, J. A., Beckwith, K. A., Natarajan, G., Woyach, J. A., Jaglowski, S., Zhong, Y., Hessler, J. D., Liu, T. M., Chang, B. Y., Larkin, K. M., Stefanovski, M. R., Chappell, D. L., Frissora, F. W., Smith, L. L., Smucker, K. A., Flynn, J. M., Jones, J. A., Andritsos, L. A., Maddocks, K., ... Byrd, J. C. (2013). Ibrutinib is an irreversible molecular inhibitor of ITK driving a Th1-selective pressure in T lymphocytes. *Blood*. <https://doi.org/10.1182/blood-2013-06-507947>
- Dupré, L., Houmadi, R., Tang, C., & Rey-Barroso, J. (2015). T lymphocyte migration: An action movie starring the actin and associated actors. *Frontiers in Immunology*, 6(NOV). <https://doi.org/10.3389/fimmu.2015.00586>
- Fagerholm, S. C., Guenther, C., Asens, M. L., Savinko, T., & Uotila, L. M. (2019). Beta2-Integins and interacting proteins in leukocyte trafficking, immune suppression, and immunodeficiency disease. *Frontiers in Immunology*. <https://doi.org/10.3389/fimmu.2019.00254>
- Ferrarini, I., Rigo, A., Montresor, A., Laudanna, C., & Vinante, F. (2019). Monocyte-to-macrophage switch reversibly impaired by Ibrutinib. *Oncotarget*. <https://doi.org/10.18632/oncotarget.26744>
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., ... Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. <https://doi.org/10.1186/gb-2004-5-10-r80>
- Hallek, M. (2019). Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. *American Journal of Hematology*.

- <https://doi.org/10.1002/ajh.25595>
- Herman, S. E. M., Mustafa, R. Z., Jones, J., Wong, D. H., Farooqui, M., & Wiestner, A. (2015). Treatment with ibrutinib inhibits BTK- and VLA-4-dependent adhesion of chronic lymphocytic leukemia cells in vivo. *Clinical Cancer Research*. <https://doi.org/10.1158/1078-0432.CCR-15-0781>
- Huttenlocher, A., & Horwitz, A. R. (2011). Integrins in cell migration. *Cold Spring Harbor Perspectives in Biology*. <https://doi.org/10.1101/cshperspect.a005074>
- Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., Carter, S. L., Cibulskis, K., Mertens, D., Sougnez, C. L., Rosenberg, M., Hess, J. M., Edelmann, J., Kless, S., Kneba, M., ... Wu, C. J. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature*. <https://doi.org/10.1038/nature15395>
- Lawrence, M., Huber, W., Pagès, H., Aboyou, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*. <https://doi.org/10.1371/journal.pcbi.1003118>
- Maffei, R., Fiorcari, S., Martinelli, S., Potenza, L., Luppi, M., & Marasca, R. (2015). Targeting neoplastic B cells and harnessing microenvironment: The “double face” of ibrutinib and idelalisib. In *Journal of Hematology and Oncology*. <https://doi.org/10.1186/s13045-015-0157-x>
- Montserrat, E., & Rozman, C. (1995). Chronic lymphocytic leukemia: Present status. In *Annals of Oncology*. <https://doi.org/10.1093/oxfordjournals.annonc.a059151>
- Olsen, T. K., & Baryawno, N. (2018). Introduction to Single-Cell RNA Sequencing. *Current Protocols in Molecular Biology*. <https://doi.org/10.1002/cpmb.57>
- Otagiri, M., Giam Chuang, V. T., Immunobiology, Murphy, K., & Weaver, C. (2017). Janeway 'S 9 Th Edition. In *America*. <https://doi.org/10.1007/978-981-10-2116-9>
- Ponader, S., Chen, S. S., Buggy, J. J., Balakrishnan, K., Gandhi, V., Wierda, W. G., Keating, M. J., O'Brien, S., Chiorazzi, N., & Burger, J. A. (2012). The Bruton tyrosine kinase inhibitor PCI-32765 thwarts chronic lymphocytic leukemia cell survival and tissue homing in vitro and in vivo. *Blood*. <https://doi.org/10.1182/blood-2011-10-386417>
- R Core team, R. F. for, & Computing., S. (2020). *R: A language and environment for statistical computing*. <https://www.r-project.org/>
- Raju, R., Gadakh, S., Gopal, P., George, B., Advani, J., Soman, S., Prasad, T. S. K., & Girijadevi, R. (2015). Differential ligand-signaling network of CCL19/CCL21-CCR7 system. *Database*. <https://doi.org/10.1093/database/bav106>
- Rendeiro, A. F., Krausgruber, T., Fortelny, N., Zhao, F., Penz, T., Farlik, M., Schuster, L. C., Nemc, A., Tasnády, S., Réti, M., Mátrai, Z., Alpár, D., Bödör, C., Schmidl, C., & Bock, C. (2020). Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nature Communications*. <https://doi.org/10.1038/s41467-019-14081-6>
- Rosati, E., Baldoni, S., De Falco, F., Del Papa, B., Dorillo, E., Rompietti, C., Albi, E., Falzetti, F., Di Ianni, M., & Sportoletti, P. (2018). NOTCH1 aberrations in Chronic lymphocytic leukemia. In *Frontiers in Oncology*. <https://doi.org/10.3389/fonc.2018.00229>
- Rossi, D., & Zlotnik, A. (2000). The Biology of Chemokines and their Receptors. *Annual Review of Immunology*. <https://doi.org/10.1146/annurev.immunol.18.1.217>
- Sauer, S. (2019). *R Markdown*. https://doi.org/10.1007/978-3-658-21587-3_26
- Sivina, M., Kreitman, R. J., Arons, E., Ravandi, F., & Burger, J. A. (2014). The bruton tyrosine kinase inhibitor ibrutinib (PCI-32765) blocks hairy cell leukaemia

- survival, proliferation and B cell receptor signalling: A new therapeutic approach. *British Journal of Haematology*. <https://doi.org/10.1111/bjh.12867>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W., Stoeckius, M., Smibert, P., & Satija, R. (2018). Comprehensive integration of single cell data. *BioRxiv*. <https://doi.org/10.1101/460147>
- Wu, C., Asokan, S. B., Berginski, M. E., Haynes, E. M., Sharpless, N. E., Griffith, J. D., Gomez, S. M., & Bear, J. E. (2012). Arp2/3 is critical for lamellipodia and response to extracellular matrix cues but is dispensable for chemotaxis. *Cell*. <https://doi.org/10.1016/j.cell.2011.12.034>
- Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. In *Genome biology*. <https://doi.org/10.1186/s13059-020-1929-3>

8. Anexos

Todas las tablas y archivos de Rmarkdown se encuentran en el repositorio GitHub
<https://github.com/javi-rey/Trabajo-fin-de-master>

A continuación, un resumen de los archivos presentes:

scRNA-seq:

- Archivo Rmd con el código utilizado

Para cada tipo celular analizado (CLL, CD4, CD8, CD14):

- 2 tablas con los genes inhibidos (treatmentup) o sobreexpresados (treatmentdown) (ejemplo: CD14_treatmentdown markers.txt)
- Una carpeta con el análisis de Enrich, conteniendo las tablas de términos GO y los gráficos asociados

ATAC-seq:

Para cada tipo celular analizado (CLL, CD4, CD8, CD14):

- Archivo Rmd con el código utilizado
- 8 tablas con los promotores de los picos alterados por Ibrutinib (ejemplo: CLL_closed0open30.txt)
- 8 tablas con los términos GO destacados a partir de las tablas anteriores (ejemplo: CD8_open0closed120enrichr.txt)
-