



**Universidad de Buenos Aires**  
Facultad de Ciencias Exactas y Naturales  
Carrera de Ciencias Biológicas

**Búsqueda de marcadores pronósticos en el clon leucémico  
y microambiente tumoral en pacientes  
argentinos con leucemia linfoblástica aguda pediátrica**


Autor: Daniel Avendaño  
Directora: Dra. María Sol Ruiz

Laboratorio de Inflamación y Cáncer  
Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales  
(IQUBICEN) - CONICET – UBA

Junio 2023



DANIEL AVENDAÑO



MARÍA SOL  
RUIZ

# Resumen

La leucemia linfoblástica aguda (LLA) es el cáncer con mayor incidencia en menores de 15 años y la principal causa de mortalidad en pediatría relacionada a cáncer. Si bien en los últimos años ha mejorado mucho la sobrevida de los pacientes, entre un 15-30% de los pacientes recaen. Actualmente se clasifica a los pacientes en tres grupos con distinto riesgo de progresión de la enfermedad, lo cual permite ajustar la intensidad del tratamiento. Esta estratificación se basa en factores genéticos, bioquímicos, clínicos y de respuesta temprana al tratamiento. El infiltrado inmunitario presente en el microambiente tumoral puede tanto promover como antagonizar el crecimiento tumoral. Actualmente no hay datos robustos sobre el poder pronóstico del infiltrado inmunitario en LLA; por lo tanto, planteamos la hipótesis de que la composición de poblaciones del sistema inmunitario en el microambiente de la médula ósea contribuye a la variabilidad en la evolución de la enfermedad y en la respuesta clínica del paciente. Este trabajo de tesis fue realizado a partir de datos de secuenciación transcriptómica obtenidos en nuestro laboratorio a partir de muestras de médula ósea de pacientes argentinos con LLA-B al momento del diagnóstico, pertenecientes al protocolo clínico multicéntrico ALL IC-GATLA 2010 (*Acute Lymphoblastic Leukemia InterContinental - Grupo Argentino de Tratamiento de la Leucemia Aguda 2010*). En primer lugar, se estimó la proporción de las distintas poblaciones inmunitarias mediante citometría digital, las cuales fueron comparadas entre pacientes con distintas características clínicas. Se utilizó la herramienta de citometría digital MIXTURE, a partir de dos firmas transcripcionales (LM22 y TIL10). MIXTURE presentó una alta reproducibilidad y evidenció gran heterogeneidad en la composición de poblaciones inmunitarias entre los pacientes al momento del diagnóstico. Se observó una mayor proporción relativa de células T CD4<sup>+</sup> (7,75% vs 3,34%, p-valor = 0,04) y una menor de monocitos (1,41% vs. 2,44%, p-valor = 0,02) en los pacientes que recayeron y/o fallecieron; mientras que los pacientes que desarrollaron toxicidad aguda relacionada al tratamiento presentaron una mayor proporción relativa de células T CD8<sup>+</sup> (TIL10: 1,69% vs. 0,83%, p-valor = 1,00x10<sup>-6</sup>; LM22: 0,97% vs. 0,48%, p-valor = 0,04) y NK (1,22% vs. 0,67%, p-valor = 0,008). No se detectaron diferencias estadísticamente significativas entre la proporción de las distintas poblaciones y las variables clínicas adicionales evaluadas en este estudio.

En segundo lugar, se estimó la abundancia de células citotóxicas mediante el cálculo de un *cytolytic score* a partir de los niveles de expresión de 5 genes de expresión específica en este tipo de células (NK y linfocitos T citotóxicos) (*GZMA*, *GZMB*, *GZMM*, *PRF1*, *GZMH*). Este índice fue estimado a partir de los datos transcriptómicos y validado por RT-qPCR, y presentó una correlación positiva significativa con la proporción de linfocitos T CD8<sup>+</sup>, T CD4<sup>+</sup> y células NK (Rho > 0,39; p-valor < 0,05). Se analizó mediante curvas de sobrevida la asociación entre la proporción de las poblaciones celulares o el valor del *cytolytic score* y el tiempo al evento muerte, recaída o muerte y/o recaída. Finalmente, mediante un análisis de todo el transcriptoma se identificó un enriquecimiento en vías biológicas relacionadas a citotoxicidad en pacientes con mayor *cytolytic score*, fortaleciendo la fiabilidad del índice calculado como así también un enriquecimiento en vías biológicas que podrían ser sugerentes de una alta inmunogenicidad en el microambiente tumoral de este grupo de pacientes.

# Identification of prognostic markers in the leukemic clone and tumor microenvironment in Argentinian pediatric patients with acute lymphoblastic leukemia.

## Abstract

Acute lymphoblastic leukemia (ALL) is the most common cancer in children and the leading cause of cancer-related mortality in pediatric patients. Although patient survival has significantly improved in recent years, 15-30% of patients experience relapse. Currently, patients are classified into three risk groups, allowing for the adjustment of treatment intensity. This stratification is based on genetic, biochemical, clinical and early treatment response factors. The immune infiltrate present in the tumor microenvironment can either promote or antagonize tumor growth. Currently, there is limited data on the prognostic value of the immune infiltrate in ALL. Therefore, we hypothesized that the composition of immune cell populations in the bone marrow microenvironment contributes to the variability in disease progression and clinical response of patients.

This thesis work was based on transcriptomic sequencing data obtained in our laboratory from bone marrow samples of B-ALL pediatric patients at the time of diagnosis, who were part of the ALL IC-GATLA 2010 (Acute Lymphoblastic Leukemia InterContinental - *Grupo Argentino de Tratamiento de la Leucemia Aguda*) multicentric clinical protocol in Argentina. First, the proportion of immune cell populations was estimated using digital cytometry and compared between patients with different clinical characteristics. We applied MIXTURE tool to perform digital cytometry, using two transcriptional signatures (LM22 and TIL10). MIXTURE showed high reproducibility and revealed significant heterogeneity in the composition of immune cell populations among patients at the time of diagnosis. A higher relative proportion of CD4<sup>+</sup> T cells (7.75% vs. 3.34%, p-value=0.04) and a lower proportion of monocytes (1.41% vs. 2.44%, p-value=0.02) were observed in patients who relapsed and/or died. On the other hand, patients who developed acute treatment-related toxicity had a higher relative proportion of CD8<sup>+</sup> T cells (TIL10: 1.69% vs. 0.83%, p-value=1.00x10<sup>-6</sup>; LM22: 0.97% vs. 0.48%, p-value=0.04) and NK cells (1.22% vs. 0.67%, p-value=0.008). No statistically significant differences were detected between the proportion of different cell populations and additional clinical variables evaluated in this study.

Secondly, the abundance of cytotoxic cells was estimated by calculating a cytolytic score based on the expression levels of 5 genes specific to these cell type (NK cells and cytotoxic T lymphocytes) (*GZMA*, *GNLY*, *GZMM*, *PRF1*, *GZMH*). This index was estimated from transcriptomic data and validated by RT-qPCR, showing a significant positive correlation with the proportion of CD8<sup>+</sup> T lymphocytes, CD4<sup>+</sup> T lymphocytes, and NK cells (Rho>0.39; p-value<0.05). The association between the proportion of cell populations or the cytolytic score value and the time to the events (death, relapse, or death and/or relapse) was analyzed by survival curve analysis. Finally, through an analysis of the transcriptome, an enrichment in cytotoxicity-related biological pathways was identified in patients with higher cytolytic scores, further supporting the reliability of the index, as well as an enrichment in biological pathways

that are suggestive of high immunogenicity in the tumor microenvironment of this group of patients.

# Agradecimientos

En primer lugar, me gustaría agradecer especialmente a mi directora, la Dra. María Sol Ruiz, quien me acompañó durante toda la tesis, quien aportó muchísimas ideas para el avance de la misma y quien siempre tuvo la paciencia de explicar y resolver hasta la más mínima duda que me surgiera. Por su pasión por la ciencia, su facilidad para expresarse y explicar cualquier tema de forma clara y amena se ha convertido en un modelo a seguir para mí.

En segundo lugar, a la Dra. Elba Vázquez, Dra. Geraldine Guerón y Dr. Javier Cotignola por aceptarme y permitir que gran parte de mi formación como biólogo haya sido en su grupo.

A los pibes del Laboratorio de Inflamación y Cáncer, Agus, Rocío, Sabrina, Flor, Sofía, Ine, Gastón, Pablo, Juan y Nacho por acompañarme y ayudarme siempre que necesite ayuda para realizar algún análisis o para resolver las dudas que se presentasen durante algún experimento. Por crear un ámbito de trabajo en el que daba gusto estar.

Me gustaría agradecer a mis padres, Mariela Haydé Miño y David Ricardo Avendaño, aunque prefiero decirles mamá y papá, y a mi hermano, Pablo, por apoyarme en todo momento y por intentar aportar ideas o soluciones siempre que se las pedía (y cuando no también). Y por haber hecho que hoy sea la persona que soy.

A mis abuelos, Elsita, Pedro y “La abuela Velia” por todas las charlas amenas que siempre me dejan con ganas de más y por haber formado una familia hermosa de la cual me siento orgulloso de pertenecer. Si en un futuro yo también formase una familia me gustaría que sea tan buena como la que ellos hicieron.

A mis tíos y primos, Felipe, María Clara, Carmelita, Gerardo, Lorena y Mercedes Miño, que a pesar de las distancias hacen que nuestra familia sea muy unida y que los sienta tan cercanos.

A mis amigos de toda la vida, Matías, Santiago, “Corti” y Caneda, por alegrar cada uno de mis días desde que los conocí. Escribiendo esto no encuentro las palabras adecuadas para describir la inmensa alegría y aporte que hicieron cada uno de ellos en mi vida.

A los “especímenes”, Emma, Luni, Lula, Vicky, Nico, “Basche” y “Chony”, y las que se sumaron durante la carrera Anita y Maru por hacer que los días de estudio en la facultad sea divertido y dieran ganas de ir a clases porque sabía que la iba a pasar bien; y también por llevar esa diversión afuera del aula y armar grupos para jugar, ver películas o simplemente “ranchar”.

# Indice

Resumen .....	1
Abstract.....	3
Agradecimientos .....	5
Indice .....	6
Abreviaturas.....	8
I) Introducción .....	9
I.1) Hematopoyesis.....	9
I.2) Leucemias.....	10
I.3) Leucemia Linfoblástica Aguda.....	11
I.4) Diagnóstico y Estratificación.....	12
I.5) Tratamiento .....	14
I.6) Estudios transcriptómicos en LLA.....	17
I.7) Microambiente Tumoral.....	17
I.8) Respuesta inmunitaria al tumor .....	19
I.9) Relevancia del problema .....	19
II) Hipótesis y Objetivos .....	21
III) Materiales y Métodos.....	22
III.1) Muestras .....	22
III.2) RNA-seq.....	22
III.3) Búsqueda de genes de referencia.....	23
III.4) Diseño de <i>primers</i> .....	23
III.5) Extracción de RNA total .....	24
III.6) Retrotranscripción .....	25
III.7) Reacción en cadena de la polimerasa (PCR) punto final.....	25
III.8) PCR cuantitativa (qPCR) .....	26
III.9) Cálculo del <i>Cytolytic Score</i> (CS).....	26
III.10) Análisis Bioinformático .....	27
III.10.1) MIXTURE.....	27
III.10.2) Gráficos .....	28
III.10.3) Análisis estadísticos.....	28
III.10.4) Análisis de agrupamiento no supervisado .....	28
III.10.5) Curvas de Kaplan Meier.....	28
III.10.6) Gene Set Variation Analysis (GSVA) .....	28

IV) Resultados .....	30
IV.1) Citometría digital .....	31
IV.1) Validación de la herramienta de deconvolución MIXTURE para la predicción de subpoblaciones leucocitarias en leucemias .....	31
IV.2) Evaluación de la asociación entre poblaciones celulares en el TME y características clínicas de los pacientes.....	38
IV.3) <i>Cytolytic Score</i> : cálculo y validación.....	47
IV.3.1) Caracterización del CS en muestras de LLA-B a partir de datos transcriptómicos (RNA-seq).....	49
IV.3.2) Ensayo de cuantificación del CS mediante RT-qPCR.....	51
IV.3.3) Búsqueda de genes de Referencia.....	52
IV.3.4) RT-qPCR de muestras de pacientes de LLA-B.....	59
IV.4) Caracterización del TME a través de vías biológicas enriquecidas a nivel transcripcional .....	61
V) Discusión.....	65
VI) Bibliografía .....	71
VII) Anexo .....	79

# Abreviaturas

ALL IC-GATLA 2010: *Acute Lymphoblastic Leukemia InterContinental - Grupo Argentino de Tratamiento de la Leucemia Aguda 2010*

CS: Índice citolítico (*cytolytic score*)

CPM: Cuentas por millón

DNA: Ácido desoxirribonucleico

EMR: Enfermedad mínima residual

GSEA: *Gene Set Enrichment Analysis*

GSVA: *Gene Set Variation Analysis*

HR: *Hazard Ratio*

LA: Leucemia aguda

LLA: Leucemia linfoblástica aguda

LLA-B: Leucemia linfoblástica aguda de tipo B

LLA-T: Leucemia linfoblástica aguda de tipo T

NK: Células asesinas naturales

RA: Riesgo alto

RE: Riesgo estándar

RI: Riesgo intermedio

RNA: Ácido ribonucleico

ROHA: Registro Oncopediátrico Hospitalario Argentino

TPM: Transcritos por millón

TCGA: *The Cancer Genome Atlas*

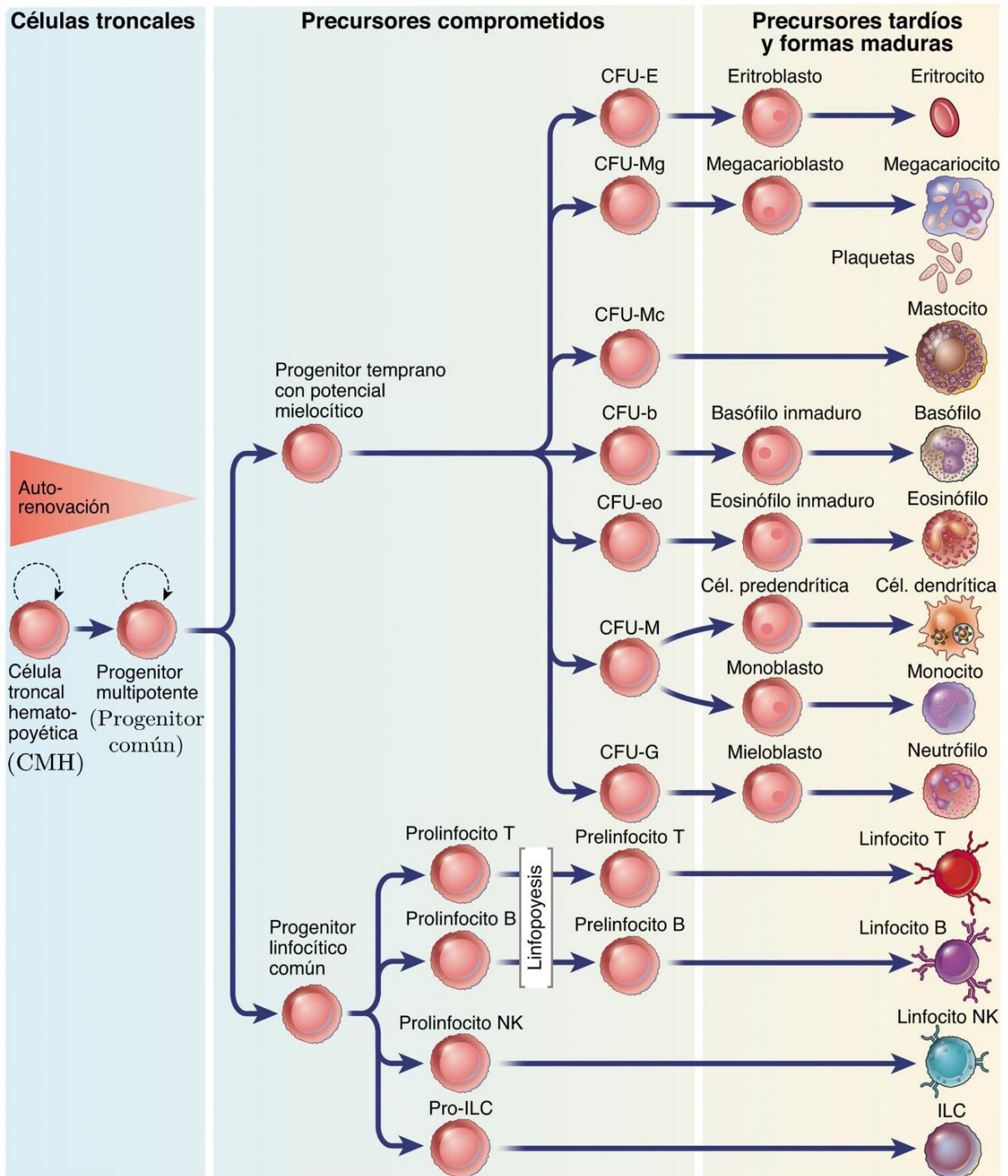


# I) Introducción

## I.1) Hematopoyesis

La hematopoyesis es el proceso mediante el cual se forman, desarrollan y maduran las distintas células que forman la sangre: eritrocitos, leucocitos y plaquetas. Este proceso se lleva a cabo a partir de una célula madre hematopoyética que es un precursor común indiferenciado. En el momento del nacimiento, la hematopoyesis se produce en todos los huesos del esqueleto, pero, con el tiempo, se restringe principalmente a la médula de los huesos planos (esternón, vértebras, huesos ilíacos y costillas). Sin embargo, en ciertas situaciones excepcionales o patológicas en las que se requiere una alta producción de células sanguíneas, el hígado y el bazo también pueden contribuir como zonas de hematopoyesis fuera de la médula ósea (Jagannathan-Bogdan & Zon, 2013).

Las células madre hematopoyéticas son capaces de diferenciarse en cualquier tipo celular dentro de un linaje sanguíneo determinado debido a su naturaleza multipotente. A partir de una célula madre hematopoyética se genera el progenitor común mielocítico-linfocítico, el cual posteriormente dará origen a los precursores independientes que se originarán los linajes mieloide y linfoide. Estas células siguen dividiéndose y madurando hasta formar las células maduras finales que conforman los elementos de la sangre periférica (Lee & Hong, 2020) (**figura 1**).



**Figura 1: Hematopoyesis: formación, desarrollo y maduración de las células sanguíneas.** CFU: unidad formadora de colonias; CFU-b: CFU de basófilos; CFU-eo: CFU de eosinófilos; CFU-G: CFU de granulocitos; CFU-M: CFU de macrófagos; CFU-Mc: CFU de mastocitos; CMH: célula madre hematopoyética; DC: célula dendrítica; ILC: células linfocíticas innatas; NK: *natural killer*. Adaptado de (Abbas, A.K., Lichtman, A.H. and Pillai, 2015).

## I.2) Leucemias

Las leucemias son un conjunto de enfermedades malignas que se caracterizan por una proliferación y diferenciación alterada de células del linaje hematopoyético. Las leucemias se

pueden clasificar, en primera instancia, en agudas y crónicas dependiendo de la rapidez que tarda en desarrollarse la enfermedad y el grado de diferenciación de células neoplásicas:

- **Leucemias agudas:** la enfermedad progresa rápidamente y se produce una acumulación de células inmaduras no funcionales en médula ósea y sangre periférica. Los pacientes con leucemias agudas suelen presentar fiebre, letargia y sangrado. También puede haber hepatoesplenomegalia, linfadenopatías y síntomas músculo-esqueléticos. En adultos pueden presentarse más síntomas relacionados con anemia. Es el cáncer más frecuente en niños/as menores de 15 años en todo el mundo (Pieters & Carroll, 2008).
- **Leucemias crónicas:** la enfermedad progresa de forma lenta y hay producción de un mayor número de células maduras que, en un principio, pueden cumplir con sus funciones fisiológicas normales. Este tipo de leucemia se presenta mayoritariamente en población adulta y muchos pacientes son asintomáticos al momento del diagnóstico, detectándose la enfermedad de manera incidental. Las leucemias crónicas pueden progresar a una leucemia aguda con el tiempo.

También se las puede clasificar de acuerdo al linaje celular afectado:

- **Leucemias mieloblásticas:** Las células afectadas pertenecen al linaje mieloide: precursores de granulocitos, eritroblastos y megacarioblastos.
- **Leucemias linfoblásticas:** Las células afectadas pertenecen al linaje linfoide: linfoblastos.

En esta tesis se estudiará en particular la Leucemia Linfoblástica Aguda (LLA).

### I.3) Leucemia Linfoblástica Aguda

Las leucemias agudas son el tipo de cáncer más común y la principal causa de mortalidad por cáncer en menores de 15 años, tanto a nivel global como en Argentina. Según el Registro Oncopediátrico Hospitalario Argentino (ROHA), las leucemias agudas representaron el 37,2% de todos los casos de cáncer diagnosticados en menores de 15 años entre 2000 y 2019, con una incidencia promedio de 503,2 nuevos casos por año, y una tasa estandarizada por edad de 49,3 casos por cada millón de niños/as (Moreno & Chaplin, 2021).

El 80,9% de las leucemias agudas corresponden a LLA, una patología que se origina en precursores linfoides del linaje B o T, definiendo así dos subtipos de LLA (LLA-B y LLA-T respectivamente). Estos precursores linfoides proliferan e infiltran la médula ósea, pudiendo comprometer diferentes órganos y/o sistemas, y causar la muerte por hemorragia y/o infección. Esta enfermedad, se caracteriza por presentar gran heterogeneidad clínica y biológica (Stanulla et al., 2020) y presenta una distribución bimodal, con un primer pico en pacientes menores de 20 años (el 80% de los casos, teniendo la mayor incidencia entre los 2 y 10 años) y el segundo a partir de los 45 años.

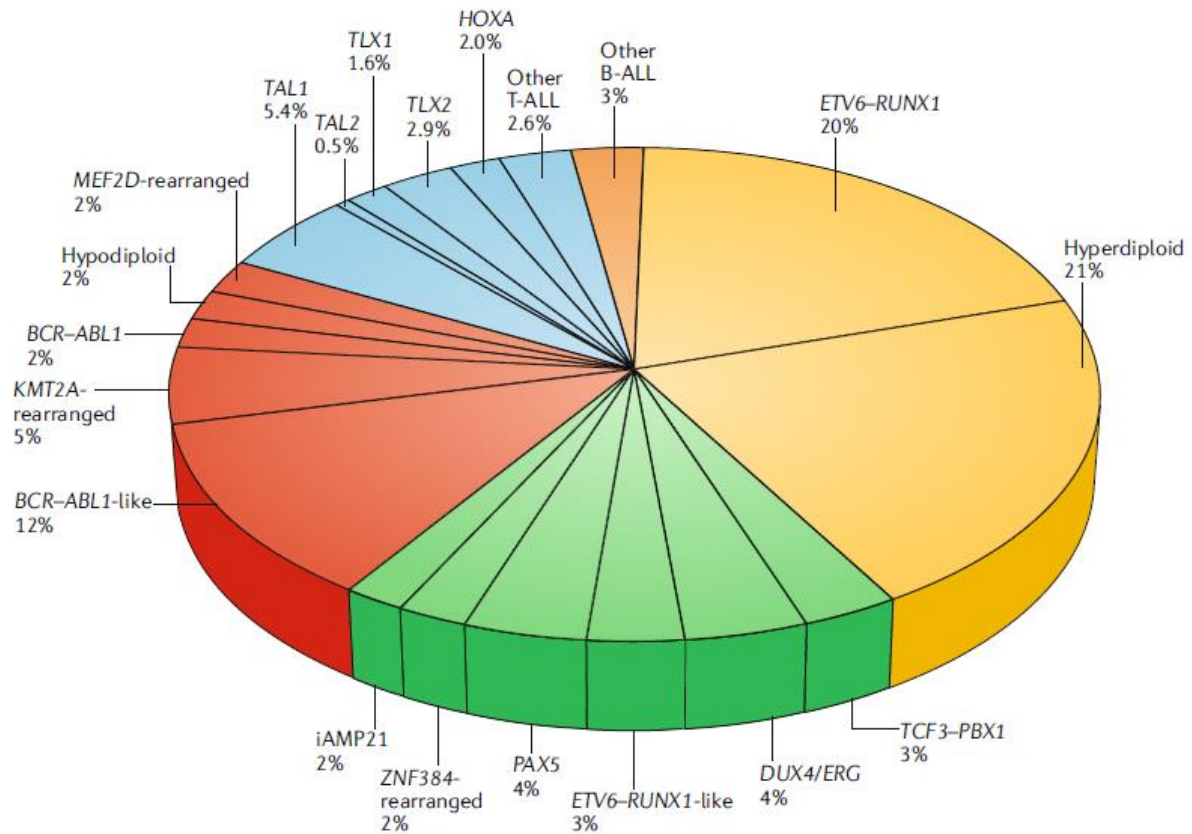
## I.4) Diagnóstico y Estratificación

Los síntomas que pueden presentar los pacientes con LLA no suelen exceder los 2 o 3 meses de evolución y pueden incluir: anemia, síndrome purpúrico-hemorrágico (lesiones rojas-púrpuras que resultan de la extravasación de sangre en la piel o mucosas), síndrome febril y dolores óseos. Sumado a la evaluación clínica, el diagnóstico se basa en la examinación citomorfológica de la médula ósea, sangre periférica y líquido cefalorraquídeo; complementado con técnicas de citogenética, biología molecular y citometría de flujo. Para el diagnóstico de LLA se utilizan las siguientes determinaciones básicas:

- **Hemograma completo y frotis de sangre periférica:** evaluación del porcentaje de blastos
- **Evaluación del líquido cefalorraquídeo:** examen fisicoquímico, citológico y citometría de flujo.
- **Punción aspiración de médula ósea** en donde se evalúa:
  - Morfología y citoquímica
  - Inmunofenotipo por citometría de flujo multiparamétrica para definir el linaje celular comprometido (precursor B, madura B, precursor T, madura T, bifeno típico)
  - Índice de ADN (estudio de la ploidía por citometría de flujo)
- **Estudios citogenéticos y moleculares:**
  - Aneuploidías: hipodiploidía (< 44 cromosomas), hiperdiploidía (>50 cromosomas)
  - Rearreglo *BCR-ABL1*, t(9:22)(q34;q11)
  - Rearreglos del gen *KMT2A*
  - Rearreglo *TCF3-PBX1*, t(1;19)(q22;p13)
  - Rearreglo *ETV6-RUNX1*, t(12;21)(q13;q22)
  - iAMP21 (amplificación intracomosomal del cromosoma 21)
  - Rearreglos del gen *IGH*
  - Perfil Phi-like: Si no hay sobreexpresión de *CRLF2* por citometría de flujo se evalúan rearreglos de genes de fusión que involucran a: *FLT3*, *IL7R*, *SH2B3*, *JAK1*, *JAK3* y *JAK2* (en combinación con genes de fusión de *CRLF2*)
  - Rearreglo *SIL-TAL1*

Las frecuencias estimadas de los diferentes subtipos descritos de LLA pediátrica se informan en la **figura 2**.

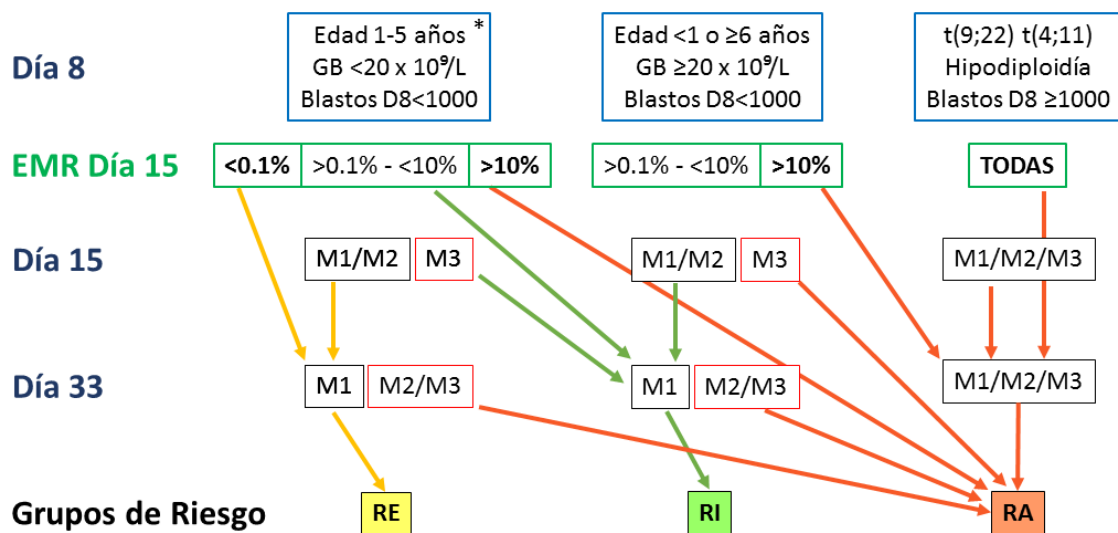
**Un número de linfoblastos > 20% en médula ósea o sangre periférica es diagnóstico de LLA.**



**Figura 2: Frecuencias estimadas de los diferentes subtipos descritos en LLA pediátrica.** Las alteraciones correspondientes a las LLA-T están coloreadas de azul. En amarillo se informan los subtipos de LLA-B de riesgo bajo, en verde los de riesgo intermedio y en rojo los de riesgo alto. Estos datos corresponden a los criterios del St. Jude Children's Research Hospital, Memphis, USA. Adaptado de (Pui et al., 2019).

Una vez que se establece el diagnóstico de LLA, los pacientes son clasificados en 3 categorías de riesgo en función de diversos factores genéticos, bioquímicos, clínicos y de respuesta temprana al tratamiento. En base a estos criterios, los pacientes son clasificados como de Riesgo Estándar (RE), Riesgo Intermedio (RI) o Riesgo Alto (RA), lo que les permite acceder a un ajuste en la intensidad del tratamiento. La enfermedad mínima residual (EMR) hace referencia al número de células leucémicas que sobreviven al tratamiento y que pueden dar lugar a la persistencia o recaída de la enfermedad. Se informa como el porcentaje de linfoblastos que se detectan en muestras de médula ósea después del tratamiento. Durante el bloque inicial del tratamiento (que se extiende desde el día 0 hasta el día 33), y según la respuesta del paciente a la quimioterapia (evaluada en función de la EMR en los días 15 y 33), los pacientes pueden permanecer en la categoría de riesgo asignada inicialmente o ser reclasificados en una categoría de mayor riesgo; sin embargo, nunca pueden ser reclasificados en una categoría de menor riesgo, aunque hayan respondido de manera óptima al tratamiento. A continuación, se presenta un esquema de la clasificación al diagnóstico detallada anteriormente (**Figura 3**). Este esquema de diagnóstico fue el utilizado dentro del protocolo conjunto del International BFM Study Group y el Grupo Argentino de Tratamiento de la Leucemia Aguda (GATLA) ALL IC-GATLA 2010/ALL IC-BFM 2009, protocolo bajo el cual fueron diagnosticados y tratados los pacientes estudiados en esta tesis.

## Diagnóstico



\* Se estudia también la alteración ETV6-RUNX1 y la hiperdiploidía. Ambas alteraciones clasifican a los pacientes como grupos de riesgo estándar

**Figura 3: Diagrama de flujo del proceso de estratificación de pacientes en grupos de riesgo estándar (RE), intermedio (RI) y alto (RA).** Al día 15 y 33 del tratamiento, según el valor de enfermedad mínima residual (EMR) se puede realizar una reclasificación, siempre hacia grupos más altos. El criterio para permanecer dentro del mismo grupo de riesgo o re-clasificarse, se esquematiza con las flechas amarilla, verde y anaranjado. M1;M2;M3 se refieren al estatus de la médula ósea: M1 <math><5\%</math> de blastos en médula ósea; M2 entre 5 y 25%; M3 >25%.

A pesar de esta categorización y el tratamiento diferencial que recibe cada grupo, todavía se observan pacientes que recaen en los tres grupos y los que presentan recurrencia por lo general fallecen (Vora et al., 2013). En un estudio de 156 pacientes pediátricos entre 2012 y 2018 recayeron en total 26 (16,7%): 10 del grupo RE y 16 del grupo de RA (Tuong et al., 2020). En otro estudio de 99 pacientes pediátricos recayeron: 3 de 38 (7,9%) de RE, 6 de 35 (17,1%) de RI y 14 de 26 (53,8%) de RA (Zhang et al., 2020). Por otro lado, según datos de la *International Agency for Research on Cancer* las leucemias corresponden al 32,5% de las muertes relacionadas a cáncer en pacientes entre 0 y 14 años, y a aproximadamente el 3,3% incluyendo pacientes de todas las edades (Globocan, 2020). Por lo tanto, **uno de los desafíos vigentes es la identificación de factores que contribuyan a explicar la variabilidad en la respuesta clínica, como así también lograr una clasificación precisa del tipo de leucemia que permita asignar a cada paciente el tratamiento más apropiado.**

## I.5) Tratamiento

El tratamiento principal para la LLA implica esquemas de poliquimioterapia que pueden ser acompañadas o no por radioterapia. Este tratamiento se divide en tres etapas: inducción, consolidación (o intensificación) y mantenimiento. Por lo general, el tratamiento dura entre 2 y 3 años, con la mayor intensidad en los primeros meses.

Esquema principal de tratamiento

**Inducción:** el objetivo de esta etapa es reducir en más del 99% la carga inicial de células leucémicas y recuperar la hematopoyesis normal. Las drogas que se usan en esta etapa son principalmente:

- glucocorticoides (prednisona, prednisolona y dexametasona)
- alcaloides de vinca como la Vincristina (inhibe la polimerización de tubulina a microtúbulos en metafase)
- otras drogas como: L-asparaginasa (descompone el aminoácido asparagina, necesario para el crecimiento tumoral), antraciclinas (producen daño a nivel del DNA), metotrexato (antimetabolito antagonista del ácido fólico) y 6-mercaptopurina (antimetabolito antagonista de la purina).
- inhibidores de tirosina quinasas como el imatinib en los pacientes con la translocación *BCR-ABL1* (*Phi+*) o perfil *Phi-like*.

La dosis y frecuencia de administración dependen del grupo de riesgo al cual pertenecen los pacientes. En esta etapa se busca que los pacientes logren una **remisión completa**, que se define de la siguiente manera:

- Médula ósea con menos del 5% de blastos, con celularidad normal o discretamente disminuida y con signos de regeneración hematopoyética.
- Ausencia de localización de infiltrados leucémicos al examen físico y/o por imágenes.
- Punción lumbar sin blastos en líquido cefalorraquídeo al día 33 del tratamiento.

**Consolidación:** comienza después de que los pacientes alcancen una remisión completa y su propósito es disminuir aún más las células leucémicas restantes en el cuerpo. Tiene una duración más prolongada que la fase de inducción y se emplean diversos enfoques según el grupo de riesgo al que pertenece el paciente. Se administran principalmente metotrexato, 6-mercaptopurina en mayores dosis y vincristina; y asparaginasa y antraciclinas en el grupo de riesgo alto. Los pacientes *Phi+* y *Phi-like* continúan también con los inhibidores de tirosina quinasa. En este momento se evalúa qué pacientes son candidatos a recibir un trasplante alogénico (pacientes de alto riesgo) y cuáles reciben otra etapa de quimioterapia más intensa durante la consolidación (conocida como reinducción).

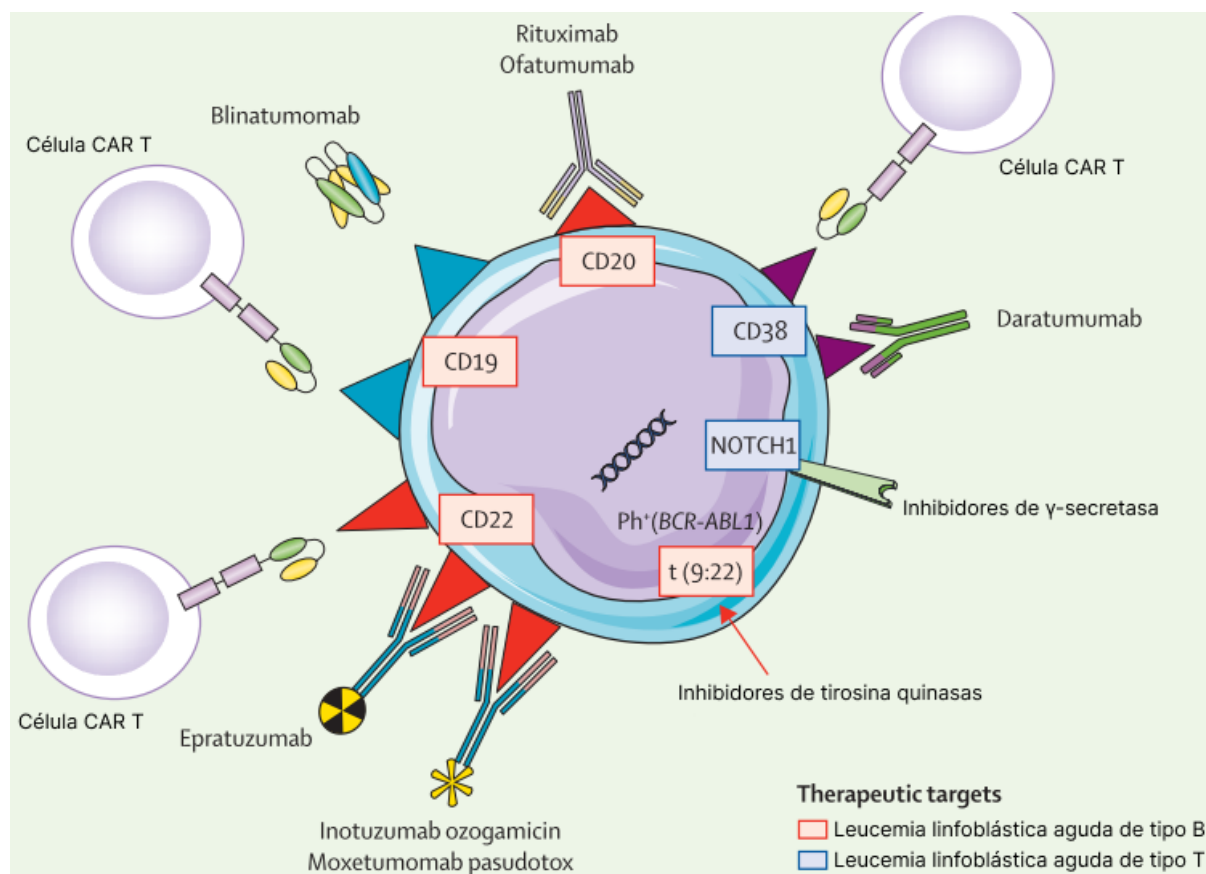
**Mantenimiento:** esta última etapa comienza si luego de las dos anteriores los pacientes continuaron en estado de remisión completa. Esta etapa puede durar 2 años o más, pero la intensidad de la quimioterapia es menor. La mayoría de los esquemas utilizan 6-mercaptopurina de forma diaria y metotrexato semanalmente.

Cerca del 30% de los pacientes presenta toxicidad relacionada al tratamiento (tromboembolismo venoso, neurotoxicidad, toxicidad ósea, pancreatitis, infecciones oportunistas, entre otras) (Schilstra et al., 2022). La intensificación de los esquemas de tratamiento está asociada a una mayor toxicidad, siendo que 2 a 3% de los pacientes fallecen por toxicidad asociada al tratamiento (Budczies, Klauschen, Sinn, Györffy, et al., 2012). Por lo tanto, para mejorar la supervivencia de los pacientes se requiere una refinación en la asignación de grupos de riesgo que permita reducir la intensidad del tratamiento en aquellos pacientes buenos respondedores y ofrecer nuevas estrategias terapéuticas a aquellos que no responden a la terapia actual.

### Inmunoterapia y terapias dirigidas

En los últimos años se han diseñado nuevas estrategias de terapias con el fin de lograr un tratamiento más dirigido y disminuir la toxicidad de la quimioterapia tradicionalmente utilizada, en especial en los grupos de riesgo alto o pacientes que presentan recaída, los cuales reciben una quimioterapia más agresiva y que tiene un límite en cuanto a la tolerancia (McNeer et al., 2020). Los mayores avances se encuentran en el campo de la inmunoterapia, con el desarrollo de anticuerpos monoclonales dirigidos hacia antígenos de superficie. Por ejemplo, el rituximab tiene como blanco a CD20, el inotuzumab ozogamicina a CD22, y el blinatumomab es dirigido contra CD3-CD19 (Jabbour et al., 2015), entre otros. En Argentina se está empezando a utilizar rituximab como parte del protocolo clínico multicéntrico ALLIC-BFM 2022 y blinatumomab como parte de las guías de tratamiento para pacientes en recaída (GATLA-LLA-REC-2022). Otra estrategia del campo de la inmunoterapia es la utilización de células T con receptores quiméricos de antígenos (CAR T-cells).

Por otro lado, la identificación de subtipos moleculares en LLA y otras leucemias ha permitido aplicar terapias dirigidas, como son los inhibidores de tirosín quinasa dirigidos a BCR-ABL1, en pacientes *Phi+* o *Phi-like*. En la **figura 4** se resumen las distintas inmunoterapias que hacen blanco a las células tumorales de LLA junto con los inhibidores de tirosinas quinasa.



**Figura 4: Esquema de las distintas inmunoterapias dirigidas a LLA, entre las cuales se incluyen anticuerpos monoclonales y CAR T-cells. Se esquematizan, además, inhibidores de tirosinas quinasa. En rojo se encuentran recuadros los distintos antígenos propios de LLA-B a los cuales se hace objetivo con anticuerpos monoclonales (CD22, CD20 y CD19) y en azul los específicos para LLA-T (CD38 y NOTCH1). Adaptado de (Malard & Mohty, 2020).**



Gracias a la optimización de protocolos de tratamiento, la identificación de distintos perfiles bioquímicos y moleculares, y al monitoreo de la respuesta temprana a la terapia mediante el seguimiento de la EMR, la tasa de supervivencia ha mejorado mucho en los últimos años, pudiendo alcanzar el 43% a los 5 años en mayores de 20 años de edad, mientras que en menores de 20 años alcanza valores de 90% en los países más desarrollados (American Cancer Society, 2023). En Argentina, la supervivencia a 5 años del periodo 2005-2014 fue de 72,2%, mientras que a los 3 años fue de 76,1% en el período 2012-2016 (Moreno & Chaplin, 2021).

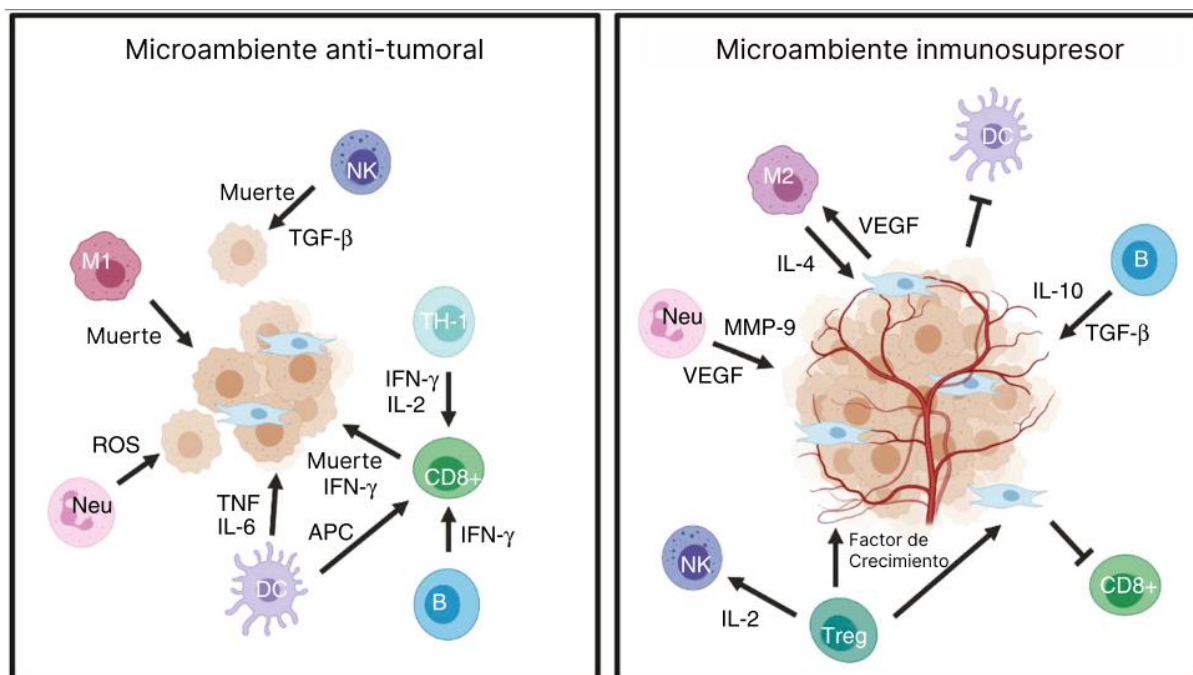
## I.6) Estudios transcriptómicos en LLA

El estudio de la expresión génica de las células tumorales ha sido muy importante en el campo de la oncología, ya que ha permitido identificar biomarcadores, subgrupos moleculares, genes y firmas génicas asociadas a respuestas clínicas diferenciales, como así también genes diana o vías de señalización que pueden servir de blancos terapéuticos.

En LLA, este tipo de estudios han permitido la identificación de subgrupos moleculares a partir de perfiles transcriptómicos similares a otros subgrupos moleculares en los que la alteración molecular ha sido identificada. Este es el ejemplo del subgrupo *Phi-like*, el cual presenta un perfil transcriptómico similar al subgrupo *Phi+*, pero en ausencia de la translocación t(9;22) (Roberts et al., 2014). Su identificación ha permitido la incorporación de inhibidores de tirosina quinasas específicos para la proteína de fusión BCR-ABL como el imatinib, dasatinib y nilotinib en el tratamiento, mejorando la respuesta terapéutica y supervivencia de estos pacientes (Malagola et al., 2016). También, mediante el estudio del transcriptoma, se han identificado distintos genes de fusión como *IGH-DUX4* o *ERG-DUX4*, los que permitieron describir nuevos subtipos de LLA-B pediátrica (Lilljebjörn et al., 2016). En el mismo trabajo se caracterizó un nuevo perfil de expresión *ETV6-RUNX1-like* que en un futuro podría permitir que los pacientes con dichas alteraciones puedan recibir un tratamiento adecuado a su perfil genético.

## I.7) Microambiente Tumoral

Se denomina microambiente tumoral (TME por sus siglas en inglés) a todos los componentes que rodean al tumor (células no tumorales, moléculas y vasos sanguíneos). Hoy en día existe amplia evidencia que hay una estrecha comunicación entre el TME y el tumor (Li et al., 2021). Esta intercomunicación permite una modulación bidireccional entre el TME y el tumor que puede tanto promover como antagonizar el crecimiento tumoral (Peña-Romero & Orenes-Piñero, 2022). Un componente muy importante del TME son las células del sistema inmunológico que, dependiendo del contexto fisiopatológico, pueden jugar un rol pro- o anti-tumoral (Anderson & Simon, 2020). En un TME anti-tumoral, se favorece una actividad citotóxica por parte de células *natural killer* (NK) y linfocitos T CD8<sup>+</sup>; los macrófagos M1 y neutrófilos también pueden intervenir. Por otro lado, se puede inducir un microambiente inmunosupresor pro-tumoral en el que se secretan citoquinas inmunosupresoras (IL10, TGF- $\beta$ , VEGF, etc.), y suelen encontrarse linfocitos con un perfil regulatorio (Tregs) o exhausto (caracterizado por no ser completamente funcionales), y los macrófagos M1 cambian su perfil a M2 inmunosupresor (Peña-Romero & Orenes-Piñero, 2022) (**Figura 5**).



**Figura 5: Esquema de interacciones descritas entre células tumorales y células inmunitarias en el microambiente tumoral.** La composición celular y citoquinas liberadas pueden resultar en un microambiente con acción predominantemente anti-tumoral (izquierda) o inmunosupresor/pro-tumoral (derecha). Adaptado de (Anderson & Simon, 2020).

Otro componente del TME son las células estromales que componen el tejido conectivo. Su composición puede variar según el tumor pero pueden incluir células del endotelio vascular, fibroblastos, adipocitos y células estrelladas (Anderson & Simon, 2020). Son relevantes para la angiogénesis, proliferación del tumor y su capacidad de hacer metástasis (Denton et al., 2018).

Por último, entre los componentes no celulares del TME, se pueden encontrar la matriz extracelular que puede funcionar como depósito de citoquinas y factores de crecimiento, y vesículas extracelulares que pueden contener lípidos, proteínas, DNA y/o RNA y que facilitan la comunicación entre células tumorales y estromales favoreciendo la inflamación, la progresión tumoral, la angiogénesis y la metástasis (Anderson & Simon, 2020).

El TME se caracteriza principalmente por la hipoxia, el pH bajo y la privación de nutrientes en comparación con el microambiente del tejido normal. La hipoxia puede inducir mutagénesis, de modo que si bien el cáncer es consecuencia de las mutaciones, las condiciones dentro del TME también pueden inducir cambios genéticos significativos en las células tumorales (Gizem Sonugür & Akbulut, 2019).

Todo lo expuesto, demuestra que el TME es complejo y cambiante en el tiempo, y depende tanto del tejido afectado como del propio tumor. El TME es un parámetro que no se utiliza para la clasificación de pacientes de LLA en los 3 grupos de riesgo mencionados previamente y ha

sido poco explorado hasta el momento en pacientes de LLA al diagnóstico y durante la respuesta al tratamiento.

## **I.8) Respuesta inmunitaria al tumor**

El sistema inmunológico se enfrenta a un desafío al tratar de reconocer y combatir las células tumorales, ya que estas son células propias del organismo. Sin embargo, existe el reconocimiento de antígenos específicos de tumores (resultado de mutaciones, deleciones, translocaciones o inserciones virales que solo están presentes en las células tumorales) y antígenos asociados a tumores (se expresan en células normales pero en las tumorales lo hacen de forma aberrante) (Feola et al., 2020). Ante una mutación, una célula normal puede sufrir una supresión intrínseca (senescencia, reparación y/o apoptosis) (Lowe et al., 2004) o bien extrínseca por medio de las respuestas inmunes innatas (células NK) o adaptativa (reconocimiento por las células dendríticas y activación de los linfocitos T citotóxicos) (Gonzalez et al., 2018). Las células citotóxicas liberan perforinas y granzimas; las primeras generan poros en la membrana de la célula tumoral y luego las granzimas disparan distintas vías de señalización induciendo a la célula blanco hacia la apoptosis (Cullen et al., 2010).

En este punto, pueden ocurrir dos escenarios: 1) la eliminación del tumor o 2) el establecimiento de un equilibrio entre el sistema inmunológico y el tumor, lo que resulta en la latencia y la edición tumoral. Entre los mecanismos de escape tumoral se encuentran: i) la secreción de citoquinas inhibitorias (por ejemplo, IL10 o TGF- $\beta$ ) o VEGF; ii) cambios en los niveles de receptores de membrana de señales pro-apoptóticas (como FasL); iii) aumento de células Tregs, macrófagos M2, macrófagos asociados a tumores, *myeloid-derived suppressor cells*; y/o iv) transición de las células efectoras a un fenotipo exhausto. Este fenotipo exhausto se caracteriza por una pérdida parcial de funcionalidad efectora, la expresión de marcadores específicos de agotamiento (TIM3, CTLA4, PD1, LAG3, entre otros), baja proliferación y cambios en su programa transcripcional (Wherry & Kurachi, 2015). Este fenotipo puede darse tanto en células T como células NK.

## **I.9) Relevancia del problema**

A pesar de todos los avances que se han mencionado en el tratamiento de la enfermedad, la LLA sigue siendo la principal causa de muerte por cáncer en pediatría. En los tres grupos de riesgo hay pacientes que recaen y este evento es uno de los principales factores de riesgo de muerte. Adicionalmente, es una enfermedad muy heterogénea en cuanto a las características genéticas del clon leucémico. Si bien se han descrito subtipos con pronóstico y terapia definida, aún existen pacientes que no pueden ser clasificados con las estrategias actuales (Gu et al., 2019), o con cariotipos o rearrreglos cromosómicos poco frecuentes, cuyo valor pronóstico es desconocido y cuya biología aún no están del todo estudiadas.

Por otro lado, el TME, como se mencionó previamente, es un parámetro poco estudiado y que podría contribuir a la eliminación del tumor, o por el contrario actuar de manera pro-tumoral, promoviendo un microambiente inmunosupresor. El TME por lo tanto tiene gran potencial

como blanco terapéutico y/o factor pronóstico. En la actualidad no se tienen en cuenta características del TME para clasificar a los pacientes de LLA en los distintos grupos de riesgo, que determinarán el tratamiento que recibirán. Por lo tanto, una caracterización del TME es un complemento prometedor para mejorar la estratificación de los pacientes e identificar nuevos blancos terapéuticos.

## II) Hipótesis y Objetivos

El perfil de alteraciones genéticas de la LLA ha revelado una gran heterogeneidad entre pacientes respecto a la expresión génica, rearrreglos cromosómicos, cambios en el número de copias, variantes de nucleótido único (tanto somáticas como germinales) y alteraciones en la ploidía. La clasificación molecular representa una estrategia prometedora para mejorar la estratificación de los pacientes, y sienta las bases para el desarrollo de terapias dirigidas, como son los inhibidores de tirosina quinasa para los pacientes *Phi+* (portadores de la translocación *BCR-ABL1*) y *Phi-like*. A pesar de estos avances, en particular en nuestro país, no hay estudios publicados que caractericen el perfil molecular de los pacientes argentinos con LLA.

El TME comprende una variedad de tipos celulares residentes e infiltrantes, factores solubles y proteínas de la matriz extracelular, que interactúan de forma recíproca con las células tumorales. Los diversos componentes del TME podrían contribuir a la heterogeneidad de la respuesta al tratamiento. Sin embargo, la clasificación actual de la LLA no tiene en cuenta factores extrínsecos a las células leucémicas.

La hipótesis del proyecto en el cual está enmarcada esta tesis es que la heterogeneidad clínica y biológica que se observa en los pacientes pediátricos con LLA, tanto al diagnóstico como durante la respuesta al tratamiento, es consecuencia de la interacción entre las características del clon leucémico y el sistema inmunitario del paciente. La elucidación de cada componente contribuiría a la estratificación de los pacientes en grupos biológicamente relevantes, con distinto riesgo de progresión y con posibilidad de acceder a tratamientos dirigidos que mejoren su eficacia y disminuyan la toxicidad asociada.

Esta tesis se focalizó en el estudio del componente inmunitario del TME en pacientes con LLA pediátrica, planteando la **hipótesis de que la composición de poblaciones del sistema inmunitario en el microambiente de la médula ósea al momento del diagnóstico contribuye a la variabilidad en la respuesta clínica.**

El **objetivo general** de este proyecto es caracterizar el componente inmunitario del microambiente tumoral en pacientes con leucemia linfoblástica aguda y evaluar su relevancia clínica. Adicionalmente se busca profundizar en el conocimiento de las interacciones entre el sistema inmunitario y las células leucémicas.

Los **objetivos específicos** son:

- 1) Identificar subpoblaciones del sistema inmunitario en el microambiente tumoral en muestras de pacientes pediátricos con LLA al diagnóstico mediante citometría digital y evaluar su asociación con características clínico patológicas de la enfermedad al momento del diagnóstico y durante la respuesta al tratamiento.
- 2) Estimar la proporción de células citotóxicas en el microambiente tumoral al momento del diagnóstico mediante un “*cytolytic score*”, optimizar su detección por RT-qPCR y evaluar su asociación con características clínico patológicas de la enfermedad al momento del diagnóstico y durante la respuesta al tratamiento.

## III) Materiales y Métodos

### III.1) Muestras

Todos los procedimientos que involucran muestras de pacientes se realizaron bajo el cumplimiento de los estándares del Comité de Ética correspondiente y en acuerdo con la Declaración de Helsinki para el trabajo con muestras humanas. Los pacientes fueron reclutados en el marco del protocolo clínico ALL IC-GATLA 2010 (<https://drive.google.com/file/d/1W5in--TICN51SSMJ2FGW82YY5A5Y1UD/view>), el cual fue aprobado por la Agencia Nacional de Medicamentos, Alimentos y Tecnología Médica. Los criterios de inclusión de los pacientes al protocolo fueron: 1) aprobación del protocolo por el Comité de ética institucional de los centros participantes, 2) LLA primaria, 3) 1-18 años de edad, 4) firma del consentimiento informado por parte de los progenitores (o tutores legales) y asentimiento informado escrito cuando fuese requerido. Los criterios de exclusión incluyeron: 1) LLA infantil (< 1 año), 2) falta de consentimiento informado, 3) LLA secundarias.

Las muestras incluidas en este trabajo provinieron de 34 pacientes atendidos en 3 hospitales de Argentina: Hospital de Niños Sor María Ludovica, La Plata, Prov. de Buenos Aires (n=8) (HLP); Hospital General de Niños R. Gutiérrez, Ciudad Autónoma de Buenos Aires (n=6) (HNG); Hospital Nacional Profesor Alejandro Posadas, El Palomar, Prov. de Buenos Aires (n=20) (POS). Las muestras correspondieron a aspirados de médula ósea (1-5 mL) de pacientes con LLA al momento del diagnóstico, previo inicio del tratamiento. Las muestras utilizadas corresponden a una alícuota de aquellas obtenidas para el diagnóstico de la enfermedad, de modo que los pacientes no fueron sometidos a una toma de muestra adicional. Los parámetros clínico-patológicos y de respuesta al tratamiento fueron registrados por los/as oncohematólogos/as derivantes.

En el caso de experimentos *in vitro* que incluyeron muestras de sangre periférica de dadores sanos (DS661 y DS662), a partir de éstas se purificaron las células mononucleares, las cuales fueron activadas *in vitro* durante 3 hs con bacterias. La activación se realizó con el objetivo de tener células citotóxicas en la muestra control para el cálculo del *cytolytic score*.

### III.2) RNA-seq

Para esta tesis se utilizaron las matrices de datos de expresión transcriptómicos generadas en nuestro grupo de investigación (Laboratorio de Inflamación y Cáncer, IQIBICEN-CONICET / Dpto. QB-FCEN, UBA), a partir de la técnica de RNA-seq, en muestras de médula ósea de pacientes del protocolo ALL IC-GATLA-2010.

Brevemente, el RNA-seq se realizó en la empresa Macrogen Corea. El procedimiento comprendió el enriquecimiento de los RNA no ribosomales con el kit Illumina Ribo-Zero Plus rRNA Depletion Kit (Illumina, California, USA), preparación de las bibliotecas y secuenciación mediante el protocolo de extremos pareados (paired-end) en equipos HiSeq2000 y HiSeq2500 Systems de Illumina. Se secuenciaron 43/79 muestras (de 39 pacientes) con

calidad y cantidad suficiente de RNA en el transcurso de 3 años en 5 lotes de secuenciación diferentes. Esta información fue luego agregada dentro de los modelos estadísticos como covariable para ajustar los resultados por una posible variabilidad inter-lote.

La calidad de los datos crudos de secuenciación fue evaluada mediante MultiQC, lo cual permitió descartar 1 muestra de baja calidad (Abbate, 2023) (Ewels et al., 2016). El número total de lecturas por muestra varió entre 28,9 y 43,3 millones de lecturas por archivo fastq. Se realizó un pseudo-alineamiento al transcriptoma de referencia humano (Homo\_sapiens.GRCh38.cdna.all.fa, obtenido de <https://www.ensembl.org/info/data/ftp/index.html>) mediante la herramienta *kallisto* (versión 0.46.1) (Bray et al., 2016). Se cuantificó la expresión a nivel de genes (anotación transcritos a genes según versión 86 de Ensembl) mediante el paquete *tximport* (Soneson et al., 2016). A partir del número de lecturas se calcularon los transcritos por millón (TPM) y cuentas por millón (CPM), y se crearon matrices de expresión transcriptómica para cada paciente. La medida TPM tiene en cuenta tanto los efectos de la longitud del gen como el número total de lecturas obtenidas durante la secuenciación de cada muestra, lo cual permite comparar la abundancia de expresión de distintos genes dentro de cada muestra. La medida de cuantificación CPM, en cambio, no tiene en cuenta la longitud del gen, pero es la requerida por métodos de normalización comúnmente utilizados para estimar expresión génica diferencial, por lo que fueron especialmente diseñados para mejorar la comparabilidad entre muestras. En esta tesis se utilizarán los datos en TPM o CPM según se indique.

### III.3) Búsqueda de genes de referencia

Se utilizó el software *NormFinder* (MOMA - Department of Molecular Medicine, Aarhus University Hospital, Denmark) en Excel (Andersen et al., 2004). Esta herramienta permite estimar la variación en la expresión de genes con el objetivo de identificar genes con baja variabilidad en sus niveles de expresión intra-grupo y entre-grupos. Para este análisis se utilizaron los datos de RNA-seq en CPM por lo que la opción de *log transform data* se dejó sin seleccionar. A su vez, en análisis separados se introdujeron distintos grupos identificadores (CS alto y bajo, recaída, muerte y toxicidad aguda asociada a tratamiento). También se seleccionó la opción *simple output only* para facilitar la interpretación.

### III.4) Diseño de *primers*

El diseño de *primers* se llevó a cabo con *Primer-BLAST* de NCBI (*National Center for Biotechnology Information*). Se estableció que los *primers* debían mapear en exones diferentes del gen de interés y con al menos un intrón de tamaño tal que el posible producto de amplificación de ADN genómico contaminante tuviera más de 1000pb y por lo tanto no amplificara en las condiciones de ciclado de la qPCR.

**Tabla 1: Características de *primers* diseñados en esta tesis\*.**

<i>Primer</i>	Secuencia de extremo 5'→3'	Tamaño producto de PCR	Temperatura de <i>Annealing</i>
<i>GNLY</i> Forward <i>GNLY</i> Reverse	GTT GAC CAA AAC ACA GGA GCT G GGG TCG CAG CAT TGG AAA CA	116 pb	60°C
<i>GZMA</i> Forward <i>GZMA</i> Reverse	AAC CAG GAA CCA TGT GCC AA GGA AAA CAC CCT CGC ACA AC	240 pb	60°C
<i>GZMH</i> Forward <i>GZMH</i> Reverse	CCT CCT GTT GGC CTT TCT TCT TCT CAC TAG GAT GCC GCC AC	151 pb	60°C
<i>GZMM</i> Forward <i>GZMM</i> Reverse	ACT GCA GAG AAA TGG CTC CCA GGA CGG GCT TGT AGC GAG	191 pb	60°C
<i>PRF1</i> Forward <i>PRF1</i> Reverse	GAC CCA CCA GGA CCA GTA CA TGC CGT AGT TGG AGA TAA GCC	179 pb	60°C
<i>NRF1</i> Forward <i>NRF1</i> Reverse	GTG CTG ATG AAG ACT CGC CT GGC CGT TTC CGT TTC TTT CC	166 pb	60°C
<i>HPS5</i> Forward <i>HPS5</i> Reverse	ATA TCC AGG CCA GAA ACC ACA AG TTC TGC AAC CAG AGT CTT TAA GGT T	230 pb	60°C
<i>PPIA</i> Forward <i>PPIA</i> Reverse	GGT ATA AAA GGG GCG GGA GG CTG CAA ACA GCT CAA AGG AGA C	156pb	60°C
<i>DYNC1LI2</i> Forward <i>DYNC1LI2</i> reverse	CTG ACC AGT GAG GAG GAG GA TTT GCC ATG CTC AGC TCC TT	171 pb	60°C

**Tabla 1:** Características de los amplicones obtenidos mediante RT-qPCR y secuencias de los *primers* utilizados.

\*Los *primers* para *PPIA* habían sido previamente diseñados en el laboratorio.

### III.5) Extracción de RNA total

Se utilizó el protocolo de *TRI Reagent*<sup>®</sup>. La concentración y pureza del RNA obtenido se midieron mediante absorbancia en un NanoDrop (ThermoFisher Scientific, Massachusetts, EE.UU.). El grado de pureza de la muestra se estimó como el cociente A260nm/A280nm (relación entre ácidos nucleicos y proteínas) y A260nm/A230nm (relación entre ácidos nucleicos y compuestos fenólicos). Consideramos un RNA con alto grado de pureza a las muestras con un valor entre 1,8 y 2 para ambas relaciones.



### III.6) Retrotranscripción

A partir de 2 µg de la muestra de RNA total se realizó una retrotranscripción (RT) siguiendo el protocolo de *TransScript® One-Step gDNA Removal and cDNA Synthesis Super Mix (Cat. No.: AT341)*, el cual utiliza tanto oligo(dT) como *random primers*. Como control de no-RT, se siguió el mismo protocolo, pero a las muestras no se les agregó la enzima RT. El control no-RT tuvo como objetivo evidenciar una contaminación de la muestra con DNA genómico.

### III.7) Reacción en cadena de la polimerasa (PCR) punto final

Para la verificación de la especificidad de los *primers* y puesta a punto, previo a la realización de las PCRs cuantitativas (qPCR) se realizaron PCRs a punto final para todos los amplicones evaluados. Las condiciones de reacción fueron:

**Tabla 2: Condiciones de reacción de PCR punto final**

Reactivo	Marca	Concentración final/Volumen
Buffer	PB-L Productos Bio-Lógicos®	1X
MgCl <sub>2</sub>	PB-L Productos Bio-Lógicos®	2 mM
dNTPs	Inbio Highway®	0.2 mM
<i>Primer Forward</i>	Integrated DNA Technologies (IDT)	0.2 µM
<i>Primer Reverse</i>	Integrated DNA Technologies (IDT)	0.2 µM
Taq Polimerasa	PB-L Productos Bio-Lógicos®	0.03 U/µl
cDNA	----	cDNA 1,8 µl
H <sub>2</sub> O	----	c.s.p 15 µl

**Tabla 2:** Condiciones de reacción de PCR punto final.

Como templado para la puesta a punto, se utilizó s cDNA proveniente de células mononucleares activadas de sangre periférica de dos donadores sanos (DS661, DS662).

Ciclado:

- 94°C por 2 min.
- 35 ciclos:
  - 92°C por 15 seg.
  - Temperatura de *Annealing* (gradiente de 56°C a 62°C) por 15 seg.
  - 72°C por 30 seg.
- 72°C por 7 min.

Se utilizó un equipo *Applied Biosystems™ Veriti™ 96-Well Thermal Cycler*. Los resultados fueron evaluados mediante electroforesis en gel de agarosa 2% teñido con bromuro de etidio (0,01 µg/ml) en buffer TAE (Tris-Acetato 40 mM; EDTA 1 mM; pH 8). Se sembraron las muestras en buffer de siembra (glicerol 50%; EDTA 50 mM; pH 8; SDS 10%; orange G 0,2%). También se sembró un marcador de peso molecular apropiado (PB-L Productos Bio-Lógicos, Argentina). La electroforesis se realizó a 100 V durante 30 min utilizando una fuente de poder (Bio-Rad, USA). Los productos de PCR se visualizaron y fotografiaron en un transiluminador de luz UV (G:Box, Syngene, Inglaterra).

### III.8) PCR cuantitativa (qPCR)

Se utilizó un termociclador *Applied Biosystems™ QuantStudio 3 Real-Time PCR Systems*. Las condiciones de reacción utilizadas fueron:

**Tabla 3: Condiciones de reacción de qPCR.**

Reactivo	Marca	Concentración/Volúmen
<i>MasterMix</i>	<i>FastStart™ Universal SYBR® Green Master (Rox)</i> (Roche)	1X
<i>Primer Forward</i>	Integrated DNA Technologies (IDT)	0,2 µM
<i>Primer Reverse</i>	Integrated DNA Technologies (IDT)	0,2 µM
cDNA (1/30)	----	1,8
H <sub>2</sub> O	----	c.s.p 15 µl

**Tabla 3:** Condiciones de reacción de qPCR. Cada muestra se analizó por triplicado.

Ciclado:

- 95°C a 10 min.
- 40 ciclos de:
  - 95°C por 20 seg.
  - Temperatura de *Annealing* por 20 seg.
  - 72°C por 30 seg (medición de fluorescencia).
- Curva de Melting:
  - 95°C por 15 seg.
  - 60°C por 1 min.
  - 60°C + 0,3°C (medición de fluorescencia) hasta 95°C

### III.9) Cálculo del *Cytolytic Score* (CS)

La estimación del CS se basa en calcular la **media geométrica** de la expresión de los genes *GNLY*, *GZMA*, *GZMH*, *GZMM* y *PRFI* de expresión específica en células citotóxicas (células NK y linfocitos T citotóxicos). Para el cálculo del CS se pueden usar tanto datos de expresión

obtenidos por RNA-seq (en unidades de TPM o CPM), o a partir de RT-qPCR (en unidades Ct). Para el caso de partir de resultados de RNAseq, el CS se calculó de la siguiente manera:

**CS = Media Geométrica de expresión de n genes** =  $\sqrt[n]{X_1 + X_2 + \dots + X_n}$  en donde  $X$  corresponde a la expresión de un gen en TPM o CPM.

El cálculo del CS a partir de datos de RT-qPCR se realizó de la siguiente manera:

- 1) Promedio de Cts de las réplicas (triplicados para las muestras de pacientes de LLA y duplicados para DS661).
- 2)  $\Delta Ct = Ct \text{ muestra} - Ct \text{ DS661}$ .
- 3)  $RQ = (1 + Ef)^{\Delta Ct}$ .
- 4) A (expresión normalizada) = RQ del gen/ RQ *NRF1* (*gen de referencia*)
- 5)  $B = \text{Log}_2(A)$
- 6) **CS = Media Geométrica de expresión de n genes** =  $\sqrt[n]{B_1 + B_2 + \dots + B_n}$ .

En donde  $Ef$  = eficiencia de amplificación por qPCR =  $[10^{(1-m)}]-1$ , y  $m$  es la pendiente resultante de graficar el Ct de una muestra en distintas diluciones seriadas al décimo, en función del logaritmo de la expresión relativa entre las diluciones.

La relativización de datos de la qPCR se basó en lo descrito por (Taylor et al., 2019)

## III.10) Análisis Bioinformático

Todos los gráficos, estimaciones de proporciones y análisis estadísticos se realizaron en el programa Rstudio versión 4.1.0.

### III.10.1) MIXTURE

La estimación de las proporciones de poblaciones celulares en las muestras de LLA se realizó con el paquete *MIXTURE* (Fernández et al., 2021) en el programa Rstudio (version de R 4.2.1, versión de Rstudio 2021.9.1.372), utilizando la siguiente línea de comando:

```
library(MIXTURE)
MIXTURE(expressionMatrix = myTPM.df,
          signatureMatrix = TIL10,
          iter = 1000,
          functionMixture = nu.svm.robust.RFE,
          useCores = 1L,
          verbose = TRUE,
          nullDist = "PopulationBased",
          fileSave = "MIXTURE_TPM_TIL10.xlsx")
```

Se utilizaron los perfiles LM22 (Newman et al., 2015a) y TIL10 (Finotello et al., 2019) para estimar la proporción de las distintas poblaciones inmunitarias.

### III.10.2) Gráficos

Para realizar los gráficos se utilizó el paquete *ggplot2* (Wickham, 2011) en Rstudio.

### III.10.3) Análisis estadísticos

Para realizar el análisis estadístico entre las proporciones estimadas por MIXTURE y las variables clínicas se utilizó un ajuste a un modelo de distribución Beta la cual es recomendada para analizar datos de proporciones. Con este fin, se utilizó el paquete *glmmTMB* (Brooks et al., 2017).

Las curvas de sobrevida se realizaron utilizando el paquete *survival* (Therneau, 2021) y la función *cutoff finder* (Budczies, Klauschen, Sinn, Gyorffy, et al., 2012).

El análisis de la concordancia por medio de test de Bland Altman (Altman, 1986) se realizó con el paquete *ggplot2* graficando en el eje Y la diferencia entre las dos variables en análisis y en el eje X el promedio entre las mismas variables. Los límites se establecen a partir de la media de la diferencia  $\pm 1.96$  del desvío estándar. El objetivo de este análisis es evaluar el grado de concordancia entre dos métodos que miden el mismo parámetro.

### III.10.4) Análisis de agrupamiento no supervisado

Se utilizó la función *ComBat* del paquete SVA (Leek et al., 2012) únicamente para este análisis con el fin de reducir la variabilidad proveniente del lote de secuenciación, ya que el agrupamiento de las muestras fue no supervisado y por lo tanto se utilizó como variable sólo la proporción relativa de cada población celular para cada muestra. Se utilizó la función *hclust* para agrupar columnas (pacientes), aclarando que se utilice *method="complete"* y que el agrupamiento se realice por correlación de *Spearman*. Los *heatmaps* se realizaron con el paquete *heatmaply* (Galili et al., 2018) y la función del mismo nombre, en donde se aclaró que el agrupamiento sea únicamente por columnas ingresando el comando *Rowv= NA*.

### III.10.5) Curvas de Kaplan Meier

Para realizar las curvas de Kaplan Meier primeramente se dividió a la cohorte de pacientes en dos grupos utilizando el método de optimización *Cutoff Finder* en Rstudio (Budczies, Klauschen, Sinn, Gyorffy, et al., 2012), en particular el método utilizado fue el llamado *Significance of correlation with survival variable*.

### III.10.6) Gene Set Variation Analysis (GSVA)

El análisis por GSVA se realizó con el paquete *gsva* (Castelo, 2023) utilizando los siguientes sets de genes descargados de *Molecular Signatures Database* (MSigDB, v2022.1):

- *H: hallmark gene sets*. (Liberzon et al., 2016)
- *C2 curated gene sets CP: Canonical pathways - Reactome subset of CP*. (Gillespie et al., 2022)
- *C5 sub-collection GO: Gene Ontology*( et al., 2023) (Ashburner et al., 2000): Esta colección de *gene sets* incluye 3 sub-colecciones que se utilizaron por separado:

- *C5 sub-collection GO: Gene Ontology - Biological Process (BP)*
- *C5 sub-collection GO: Gene Ontology - Cellular Component (CC)*
- *C5 sub-collection GO: Gene Ontology - Molecular Function (MF)*
- *C5 sub-collection HPO: Human Phenotype Ontology (Köhler et al., 2021)*
- *C7: immunologic signature gene sets - ImmuneSigDB subset of C7 (Godec et al., 2016)*

Los parámetros introducidos en las líneas de comando fueron los siguientes:

```
library(GSVA)
gsva(myTPM.filt.35m, GeneSet_C7_immune.SigDB,
      method=c("gsva"),
      kcdf=c("Gaussian"),
      abs.ranking=FALSE,
      min.sz=10,
      max.sz=500,
      parallel.sz=1L,
      mx.diff=TRUE,
      tau=switch("gsva",gsva=1, ssgsea=0.25,NA),
      ssgsea.norm=TRUE,
      verbose=TRUE)
```

Por otro lado, los contrastes para la comparación del Enrichment Score entre grupos (que en nuestro caso eran CS alto y bajo) se realizó utilizando como covariables el lote de secuenciación, el sexo y el grupo de riesgo al que pertenecían los pacientes.

## IV) Resultados

Las muestras utilizadas en este trabajo de tesis fueron reclutadas como parte de un proyecto de investigación más amplio dentro del protocolo clínico ALLIC GATLA 2010. Todas las muestras fueron obtenidas al momento del diagnóstico y fueron alícuotas de las muestras utilizadas para hacer el diagnóstico y tipificación de la LLA. En la **Tabla 4** se resumen las características clínicas de los pacientes. Dado que las muestras POS008 y HNG004 serían descartadas en próximos análisis y no se tuvieron en cuenta en los resultados de la tesis, no fueron incluidas en la **Tabla 4**.

**Tabla 4: Características clínicas de las muestras utilizadas en la tesis.**

<b>Número de muestras</b>	32
<b>Edad al diagnóstico<sup>1</sup></b>	7 años (2-18)
<b>Tiempo de seguimiento<sup>1</sup></b>	32 meses (2-52)
<b>Subtipo LLA</b>	<u>Común</u> : 26 (81%)
	<u>pre-B</u> : 6 (19%)
<b>Respuesta a prednisona<sup>2</sup></b>	<u>Buena</u> : 28 (88%)
	<u>Pobre</u> : 4 (12%)
<b>Fallece</b>	<u>Sí</u> : 5 (16%)
	<u>No</u> : 27 (84%)
<b>Recaída</b>	<u>Sí</u> : 5 (16%)
	<u>No</u> : 27 (84%)
<b>Grupo de riesgo</b>	<u>Estándar</u> : 3 (9,375%)
	<u>Intermedio</u> : 22 (68,75%)
	<u>Alto</u> : 7 (21,875%)
<b>EMRd15Cat<sup>3</sup></b>	<u>&lt; 0.1%</u> : 11 (34,375%)
	<u>1-10%</u> : 14 (43,75%)
	<u>≥ 10%</u> : 5 (15,625%)
	<u>ND<sup>4</sup></u> : 2 (6,25%)
<b>Toxicidad aguda</b>	<u>Sí</u> : 9 (28,165%)
	<u>No</u> : 22 (68,75%)
	<u>ND<sup>4</sup></u> : 1 (3,125%)

**Tabla 4: Características clínicas de las muestras utilizadas en la tesis.** <sup>1</sup> Mediana (rango); <sup>2</sup> Buena respuesta: <1000 blastos/ $\mu$ l al día 8 del tratamiento; <sup>3</sup> Enfermedad Mínima Residual al día 15: % de blastos en médula ósea. <sup>4</sup> No determinado.

## IV.1) Citometría digital

Una técnica que ha revolucionado diversos campos como la inmunología, biología molecular, biología del cáncer y vigilancia de enfermedades infecciosas, es la citometría de flujo, una herramienta que permite identificar determinados marcadores dentro de una muestra analizando célula por célula la presencia o no de dicho marcador. Esto permite una caracterización muy detallada de las poblaciones celulares que componen una muestra biológica. Esta técnica es particularmente útil para caracterizar las poblaciones del sistema inmunitario.

En los últimos años se han desarrollado diversos métodos de deconvolución que permiten realizar lo que se conoce como citometría digital. Se basa en comparar el perfil de expresión de genes de una muestra compleja (es decir, con mezcla de poblaciones celulares) a una firma transcriptómica específica de distintos tipos celulares. De este modo, se obtienen proporciones estimadas de cada tipo celular. Esto permite aprovechar la disponibilidad de una gran cantidad de transcriptomas de tumores secuenciados y depositados en repositorios de acceso público, y así explorar el componente inmunitario y estromal de los mismos. Es a su vez de particular utilidad para aquellas muestras en cantidad limitante o difícil procesamiento para citometría de flujo convencional.

Se han desarrollado muchas herramientas para realizar citometría digital con distintos métodos de deconvolución. Dentro de los más utilizados se encuentran el método diseñado por Abbas et al. en 2009 (Abbas et al., 2009), CIBERSORT (Newman et al., 2015b), CIBERSORTx (Newman et al., 2019), quanTIseq (Finotello et al., 2019) y ABIS (Monaco et al., 2019).

## IV.1) Validación de la herramienta de deconvolución MIXTURE para la predicción de subpoblaciones leucocitarias en leucemias

Con el objetivo de caracterizar el componente inmunitario del microambiente tumoral en las muestras de pacientes pediátricos con LLA se utilizó la herramienta MIXTURE (Fernández et al., 2021) para estimar la abundancia de distintas poblaciones celulares del sistema inmunitario en la médula ósea al momento del diagnóstico. Se escogió esta herramienta de deconvolución para el análisis ya que, según sus autores, presenta mayor robustez en la estimación de las proporciones de poblaciones celulares que ABBAS, ABIS, quanTIseq y CIBERSORT, y porque fue desarrollada en nuestro país. Por otro lado, tanto MIXTURE como el resto de las herramientas de deconvolución con las cuales fue validada, fueron testeados únicamente en tumores sólidos, con lo cual decidimos evaluar su desempeño en el caso de las leucemias, donde las células tumorales comparten un origen común en el desarrollo con las células del sistema inmunitario, y además siendo muestras de médula ósea, se combinan poblaciones

inmunitarias residentes con aquellas asociadas a la presencia de células leucémicas. Se introdujeron como input los datos de RNA-seq en TPM y se utilizaron 2 perfiles transcriptómicos, LM22 y TIL10, que distinguen 22 y 10 poblaciones celulares inmunitarias, respectivamente.

LM22 es una matriz de perfiles de expresión compuesta por 547 genes que distinguen 22 poblaciones celulares hematopoyéticas. Para generarla los autores obtuvieron datos de expresión génica de dominio público para 22 subconjuntos de leucocitos, filtraron los genes de expresión diferencial específicos para cada población inmunitaria y luego la validaron por medio de CIBERSORT con datos externos de leucocitos purificados (Newman et al., 2015a). Las poblaciones celulares que distingue la matriz resultante son:

- *B cells naive*
- *B cells memory*
- *Plasma cells*
- *T cells CD8<sup>+</sup>*
- *T cells CD4<sup>+</sup> naive*
- *T cells CD4<sup>+</sup> memory resting*
- *T cells CD4<sup>+</sup> memory activated*
- *T cells follicular helper*
- *T cells regulatory (Tregs)*
- *T cells gamma delta*
- *NK cells resting*
- *NK cells activated*
- *Monocytes*
- *Macrophages M0*
- *Macrophages M1*
- *Macrophages M2*
- *Dendritic cells resting*
- *Dendritic cells activated*
- *Mast cells resting*
- *Mast cells activated*
- *Eosinophils*
- *Neutrophils*

La matriz TIL10 se generó a partir de un compendio de datos de RNA-seq consistente en 19.423 genes y 53 bibliotecas de células del sistema hematopoyético y tumorales. Se filtraron los genes que tuviesen una expresión diferencial en una población celular con respecto a las demás, y ausentes en las células tumorales analizadas, llegando a una matriz final con 170 genes capaz de distinguir 10 poblaciones inmunitarias (Finotello et al., 2019):

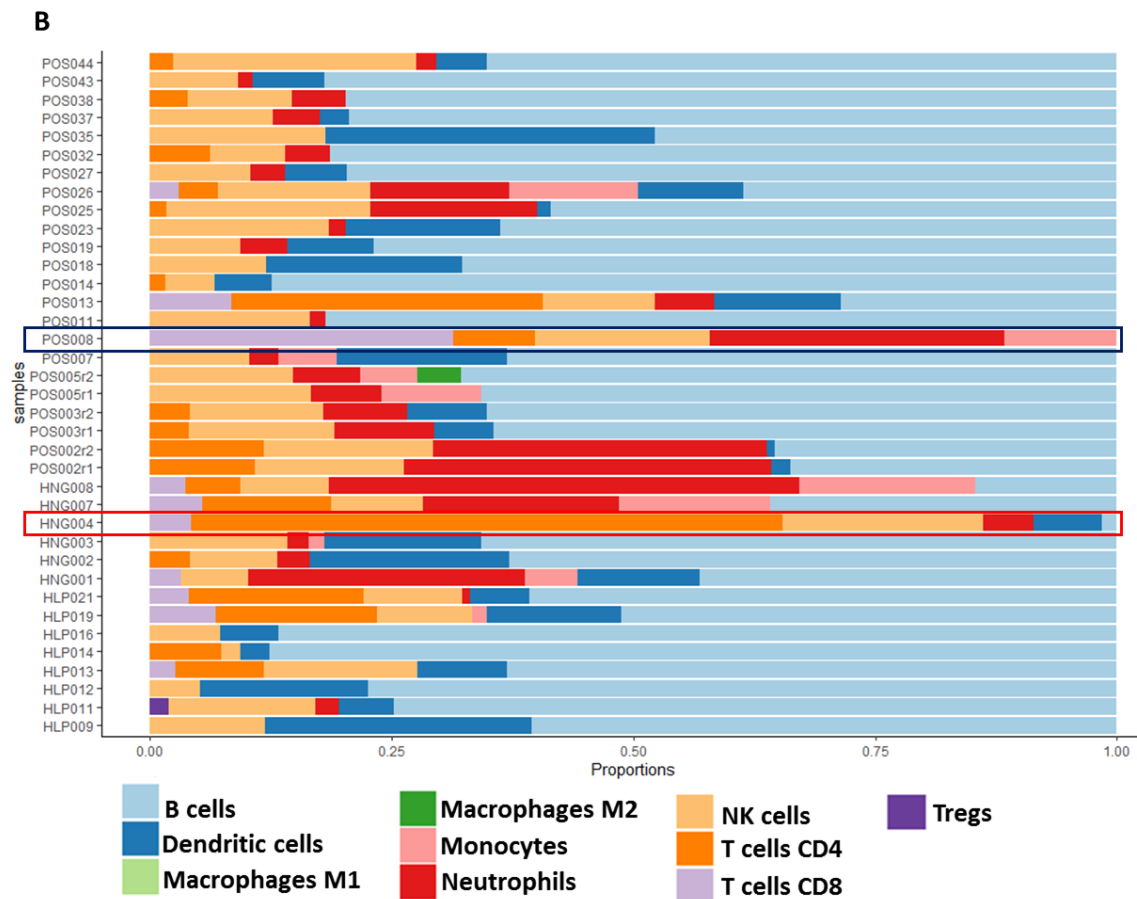
- *B cells*
- *Macrophages M1*
- *Macrophages M2*
- *Monocytes*



- *Neutrophils*
- *NK cells*
- *T cells CD4*
- *T cells CD8*
- *Tregs*
- *Dendritic cells*

Mediante el análisis de RNA-seq de las muestras de los pacientes reclutados en nuestro protocolo, pudimos estimar por MIXTURE la abundancia de las poblaciones inmunitarias según los perfiles LM22 y TIL10. Los resultados se muestran en la **figura 6**.





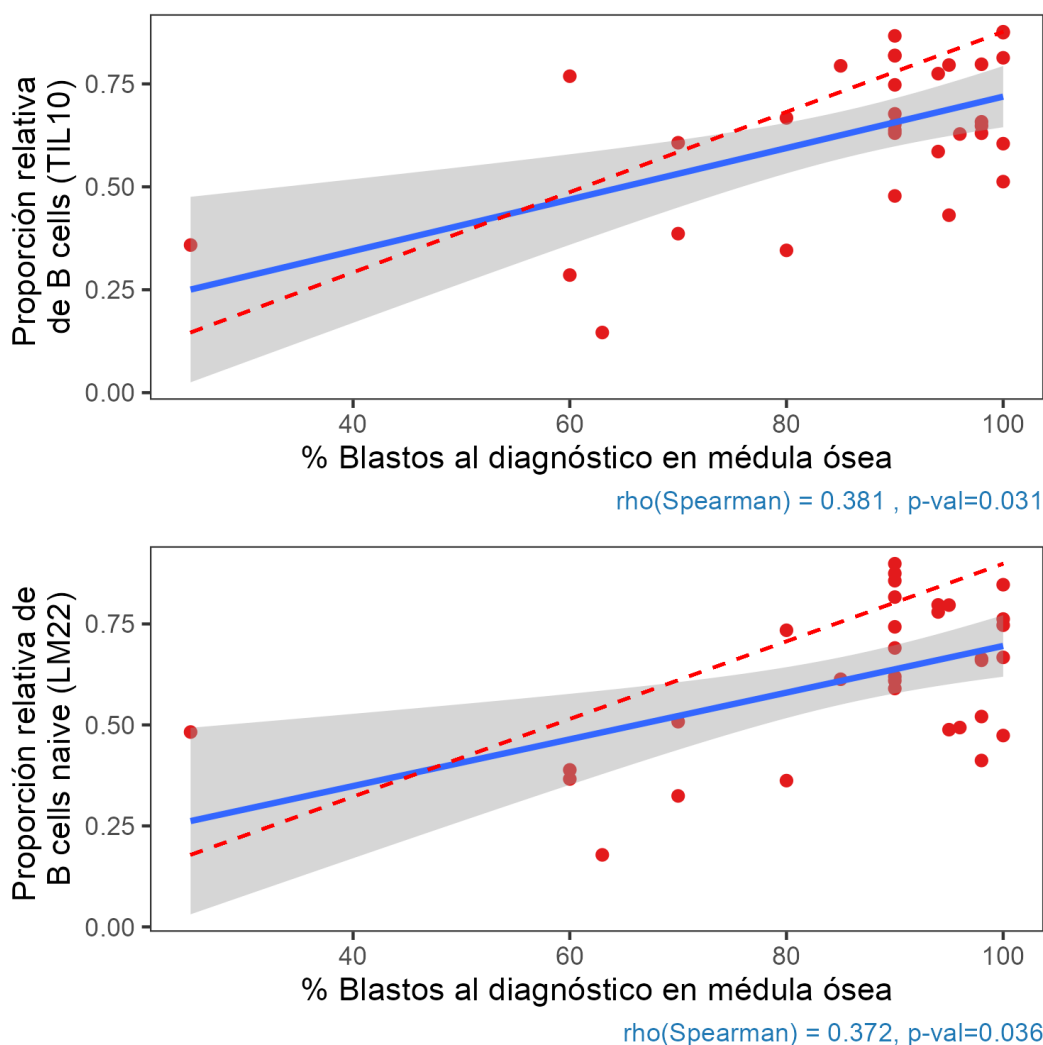
**Figura 6: Proporciones de poblaciones inmunitarias estimadas por MIXTURE en muestras individuales.** Se utilizaron las firmas transcriptómicas LM22 (A) y TIL10 (B). Recuadrada en rojo se indica la única muestra de tipo LLA-T (HNG004). Recuadrada en azul se indica una muestra de subtipo atípico (POS008) según la estimación de MIXTURE. Las muestras restantes son de tipo LLA-B.

El análisis por MIXTURE logró identificar perfiles con predominio de células de linaje B en las LLA-B, y con predominio de células de linaje T en la LLA-T (**figura 6**). Un punto importante a considerar en este análisis es que los perfiles utilizados (LM22 y TIL10) no están diseñados para la identificación de células precursoras (blastos) y que, además, MIXTURE fue desarrollada para tumores sólidos; por lo tanto, las proporciones de linfocitos B maduros y linfocitos T podrían estar sobreestimadas debido a los blastos B o T presentes en las LLA-B y LLA-T, respectivamente. Además, en el RNA-seq se incluyeron 3 muestras por duplicado para evaluar la variabilidad en la secuenciación. El análisis de los duplicados demostró resultados comparables entre ambas réplicas (POS002r1/POS002r2, POS003r1/POS003r2, POS005r1/POS005r2; **tabla 5**), siendo la mayoría de los residuos menores a  $|0,01|$  (**figura 7**). Las únicas excepciones fueron la aparición de una población de macrófagos M2 en una de las réplicas de POS005 y no en la otra (tanto en TIL10 como en LM22) y para POS003 LM22 se ve que una de las réplicas tiene linfocitos T CD4<sup>+</sup> *naive* (y no linfocitos T CD4<sup>+</sup> *memory resting*) mientras que en la otra réplica pasa lo opuesto con proporciones muy similares. Esto último sugiere que la identificación de estados funcionales sería menos reproducible dado que comparten la expresión de aquellos genes que le otorgan identidad a la población celular



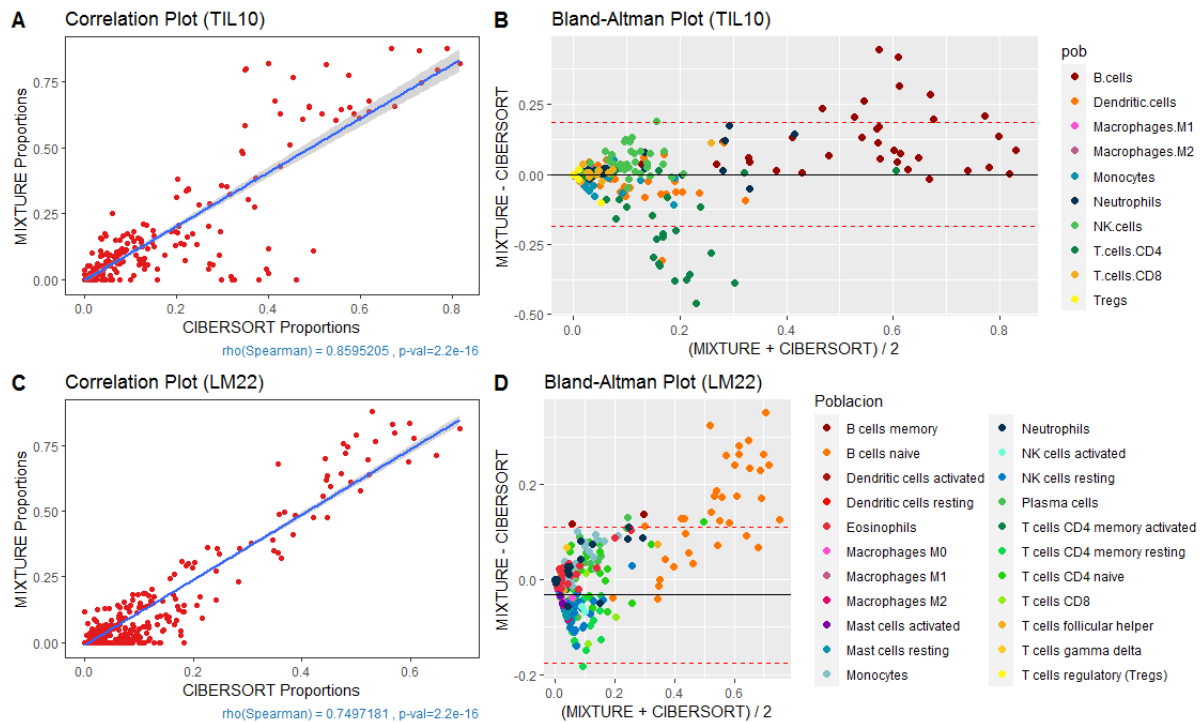
entre ambas era alta. Para ello se sumaron las distintas poblaciones pertenecientes a un mismo linaje que estimaba LM22. Las poblaciones que presentaron una correlación positiva significativa fueron: células B (Rho Spearman = 0,75; p-val =  $7,3 \times 10^{-8}$ ), Monocitos (Rho Spearman = 0,345; p-val = 0,036), Neutrófilos (Rho Spearman = 0,8, p-val =  $2,5 \times 10^{-9}$ ), Natural Killer (Rho Spearman = 0,44, p-val = 0,006) y linfocitos T CD8<sup>+</sup> (Rho Spearman = 0,45; p-val = 0,005). Los Macrófagos M2 (Rho Spearman = -0,028; p-val = 0,87), linfocitos T CD4<sup>+</sup> (Rho Spearman = 0,31; p-val = 0,063) y T regulatorios (Tregs) (Rho Spearman = -0,028; p-val = 0,87) no presentaron una correlación significativa. Por último, para las poblaciones de Macrófagos M1 y células Dendríticas no se pudo calcular una correlación debido a que las proporciones eran iguales a cero para todas las muestras en alguna de las 2 matrices. Las diferencias detectadas son esperables debido a que ambas matrices comparten una baja proporción de genes (48 genes en común): el porcentaje de solapamiento con TIL10 de LM22 es de 8,78% (48/547), mientras que el porcentaje de solapamiento con LM22 de TIL10 es de 28,23% (48/170). Es debido a esto que se decidió proceder con el análisis de ambas matrices en paralelo.

Con el objetivo de validar, al menos parcialmente, la herramienta MIXTURE, se compararon las proporciones de células B estimadas por ésta con el porcentaje de blastos medidos al diagnóstico por citomorfología (**figura 8**). Dado que las matrices LM22 y TIL10 no incluyen como población individual a los precursores (blastos), realizamos esta comparación bajo el supuesto de que la población identificada como células B sirve como indicador de células precursoras y maduras de linaje B. Se observó una correlación positiva en ambos casos (LM22: rho Spearman = 0,372, p-valor < 0,05; TIL10: rho Spearman = 0,381, p-valor < 0,05). Se observa que la línea de regresión (línea continua azul) está por debajo de la línea de identidad (punteada roja), lo cual es esperable por el menor rango dinámico de la citomorfología.



**Figura 8: Comparación de MIXTURE con citomorfología.** Gráfico de correlación entre las proporciones de las poblaciones de linaje B estimadas por MIXTURE y el porcentaje de blastos medidos al diagnóstico en médula ósea por citomorfología. La línea punteada roja indica la identidad.

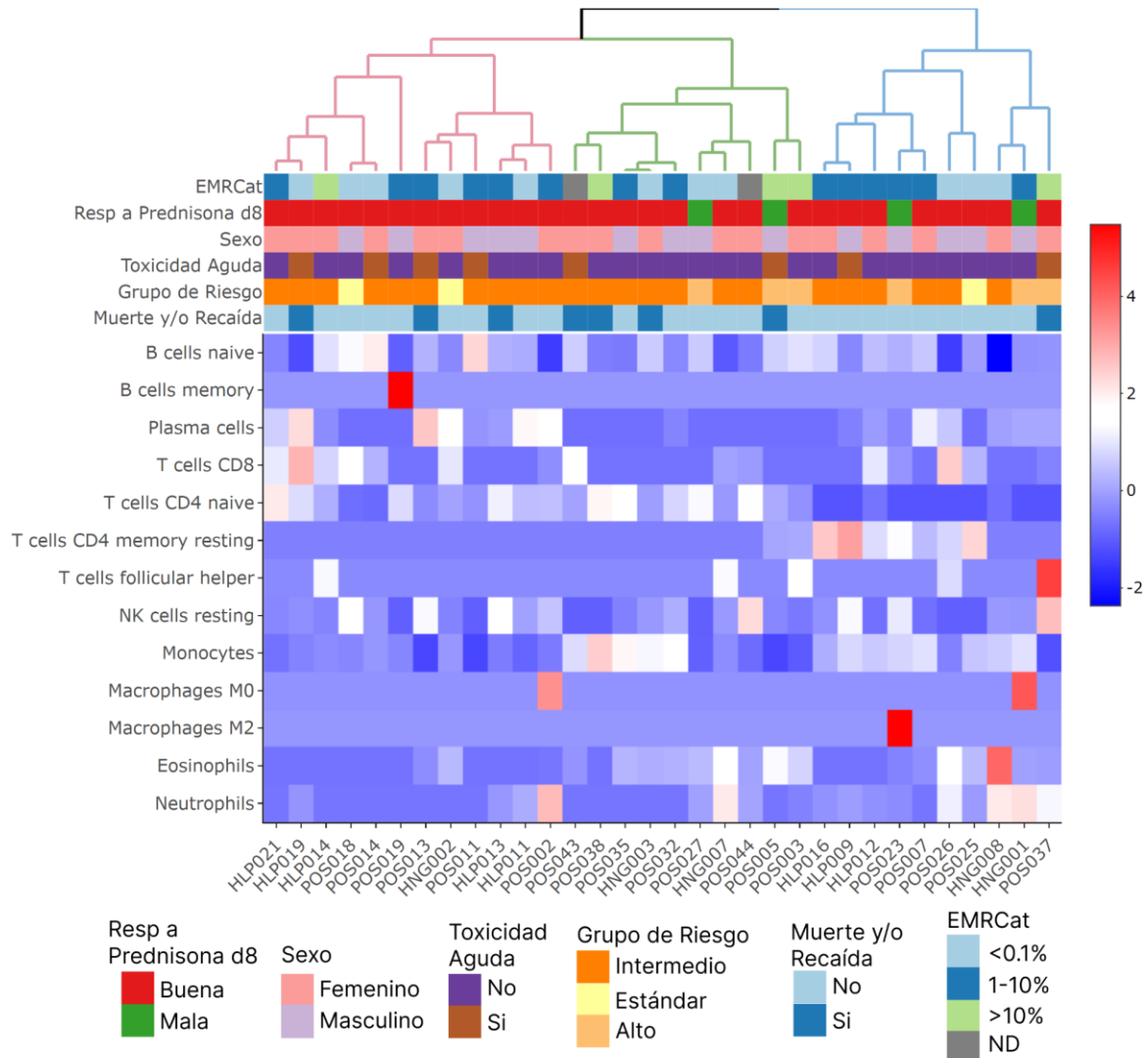
Por otro lado, se comparó la estimación de proporción de poblaciones utilizando las mismas matrices con la herramienta de deconvolución CIBERSORT. Se observó una correlación positiva entre las proporciones estimadas por MIXTURE en comparación a las proporciones estimadas por CIBERSORT con ambas matrices (TIL10:  $\text{Rho Spearman} = 0,859$ ,  $p\text{-valor} = 2,2 \times 10^{-16}$ , **figura 9 A**; LM22:  $\text{Rho Spearman} = 0,749$ ,  $p\text{-valor} = 2,2 \times 10^{-16}$ , **figura 9 C**). Mediante el análisis de Bland Altman, para LM22 (**figura 9 B**) se observó una concordancia del 95,7% (779/814) y para TIL10 (**figura 9 D**) la concordancia fue de 93% (344/370). En el caso de TIL10, el método de MIXTURE y CIBERSORT discreparon en la estimación de células B y células T CD4<sup>+</sup>; mientras que con LM22 las discrepancias se detectaron para la estimación de células B vírgenes (*B cells naive*).



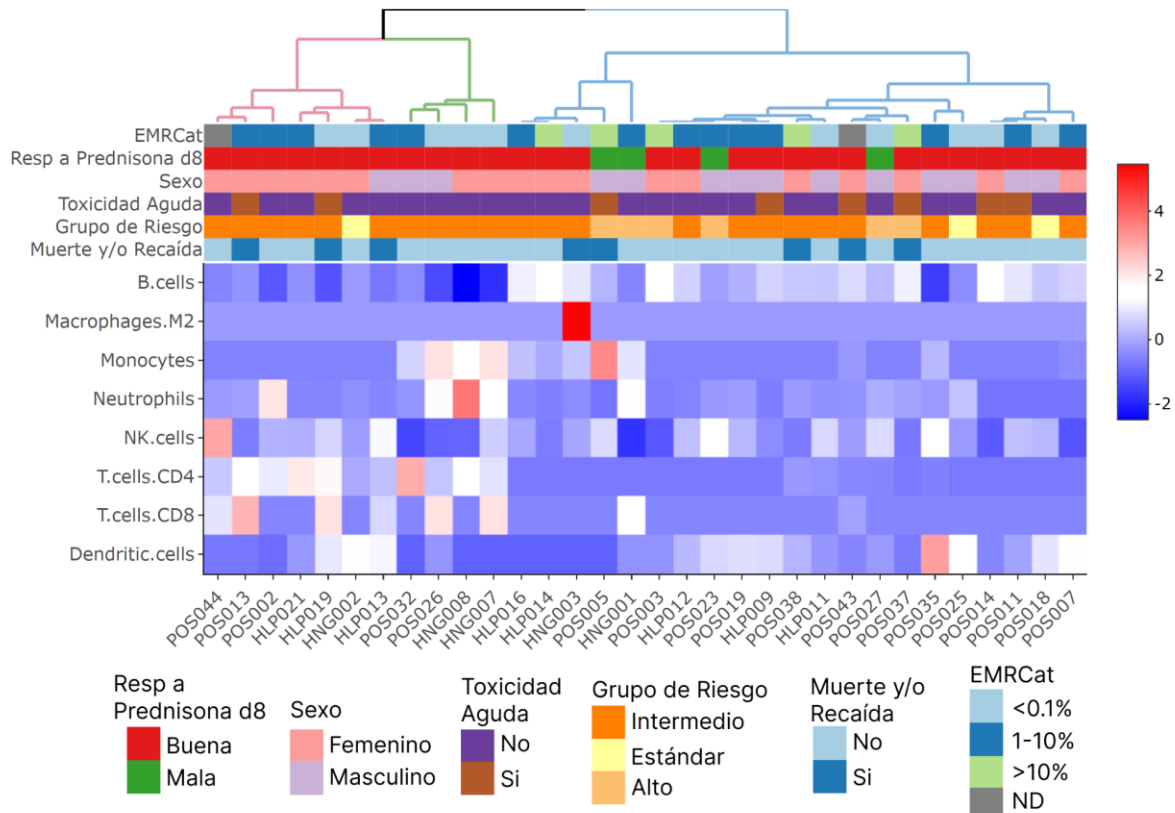
**Figura 9: Comparación entre las proporciones celulares estimadas por MIXTURE y CIBERSORT. A:** gráfico de correlación entre MIXTURE (eje y) y CIBERSORT (eje x) utilizando la matriz de TIL10. **B:** análisis de Bland Altman para la firma TIL10. Eje y: diferencia entre proporciones de MIXTURE y las de CIBERSORT. Eje x: promedio entre ambas herramientas de deconvolución. En línea punteada se indica la media de la diferencia entre las proporciones estimadas por las dos herramientas  $\pm 1,96$  veces el desvío estándar. **C:** idem A pero con la firma LM22. **D:** idem B pero con la firma LM22.

## IV.2) Evaluación de la asociación entre poblaciones celulares en el TME y características clínicas de los pacientes

A continuación, se realizó un agrupamiento no supervisado de las muestras según las proporciones de las distintas poblaciones con el objetivo de identificar conjuntos de pacientes con un perfil similar en su composición de poblaciones inmunitarias, como así también evaluar si estos agrupamientos podían asociarse a las características clínicas. Para ello se eliminaron las poblaciones celulares que tuviesen una proporción igual a cero para todas las muestras ya que no aportaban información (**figura 10** y **figura 11**), y se utilizaron las proporciones relativas obtenidas luego de una corrección por efecto de lote de secuenciación (ver Materiales y métodos). Se definieron 3 grupos de muestras (*clusters*) a partir del resultado del agrupamiento no supervisado.



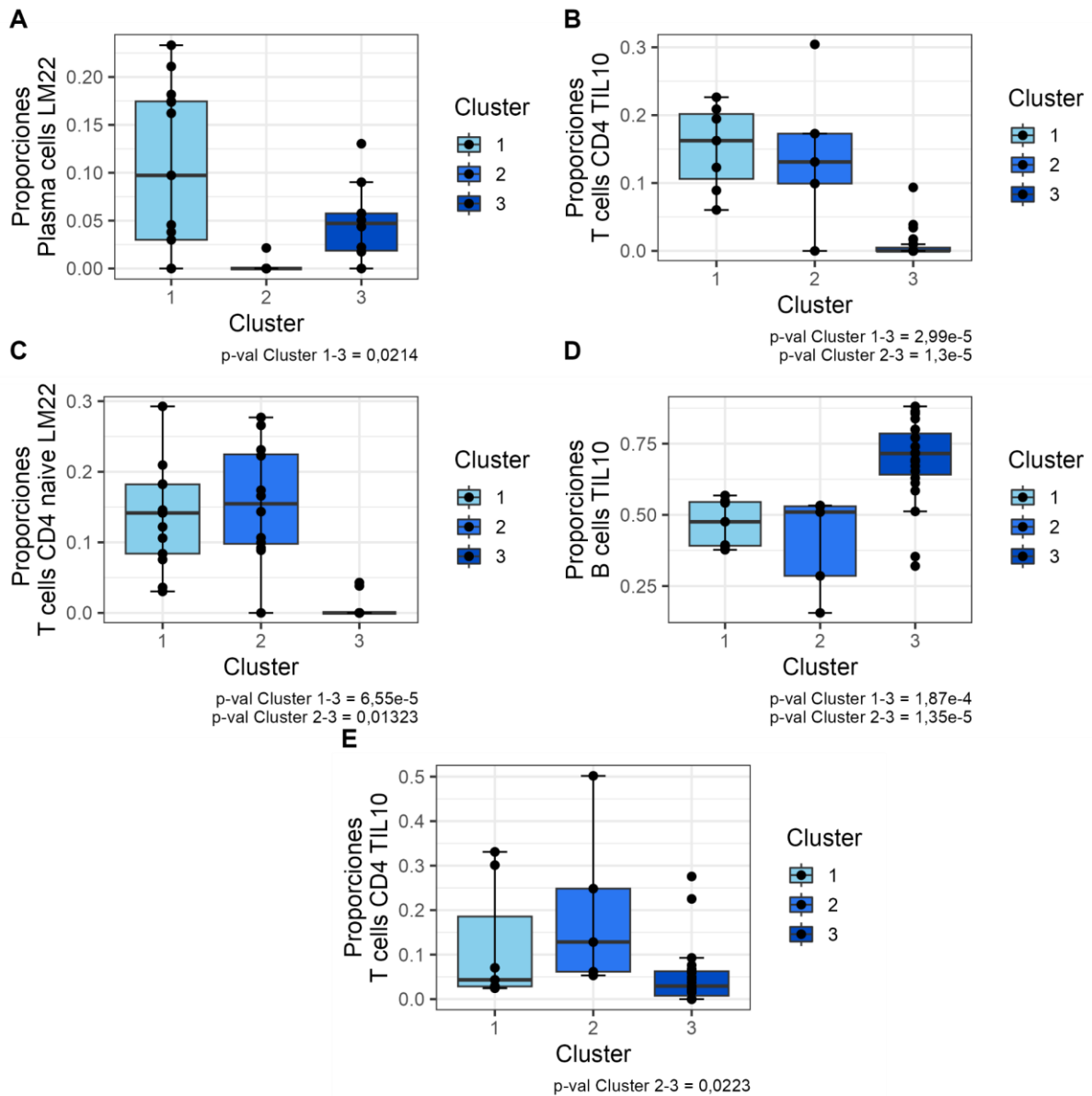
**Figura 10: Análisis de agrupamiento no supervisado de las muestras de LLA según la composición de poblaciones inmunitarias.** Los colores del *heatmap* representan las proporciones relativas de cada población celular estimada por MIXTURE (LM22), y las categorías correspondientes en cada característica clínica. Resp a Prednisona d8: respuesta al tratamiento con prednisona al día 8; Toxicidad Aguda: toxicidad aguda relacionada al tratamiento; EMRCat: Categoría de Enfermedad Mínima Residual al día 15 en base al porcentaje de blastos en médula ósea.



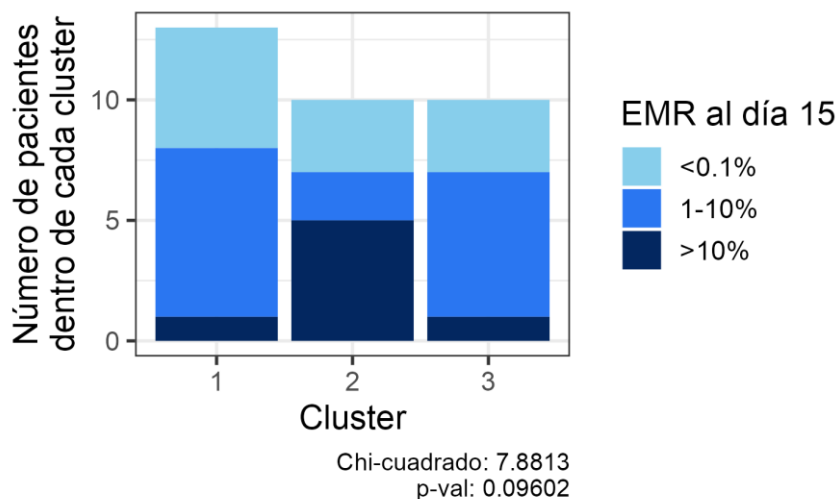
**Figura 11: Análisis de agrupamiento no supervisado de las muestras de LLA según la composición de poblaciones inmunitarias.** Los colores del *heatmap* representan las proporciones relativas de cada población celular estimadas por MIXTURE (TIL10), y las categorías correspondientes en cada característica clínica. Resp a Prednisona d8: respuesta al tratamiento con prednisona al día 8; Toxicidad Aguda: toxicidad aguda relacionada al tratamiento; EMRCat: Categoría de Enfermedad Mínima Residual al día 15 en base al porcentaje de blastos en médula ósea.

Se observó una concordancia parcial en el agrupamiento de pacientes para LM22 y TIL10. Para TIL10, el cluster 3 presentó diferencias significativas en las proporciones de células B, células T CD4<sup>+</sup> respecto a los clusters 1 y 2 y neutrófilos entre los clusters 2 y 3 ( $p < 0,05$ ; modelo: distribución beta) (**Figura 12, B, D y E, tabla S.1**). Para LM22, se detectaron diferencias significativas entre *clusters* en las proporciones de células plasmáticas y células T CD4<sup>+</sup> *naive* ( $p < 0,05$ ; modelo: distribución beta) (**Figura 12 A y C, tabla S.1**). Sin embargo, no se detectaron diferencias estadísticamente significativas entre clusters respecto a las variables clínicas muerte/recaída, grupo de riesgo, respuesta a la prednisona, toxicidad aguda y EMR ( $p > 0,05$ ; prueba de Chi-cuadrado). Sin embargo, sí se observó una diferencia marginal en cuanto a la distribución de pacientes pertenecientes a distintas categorías de EMR entre *clusters* utilizando la firma LM22 (Chi-cuadrado = 7,88,  $p\text{-val} = 0,096$ ), resaltando el mayor número de pacientes con EMR > 10% (peor respuesta al tratamiento) en el cluster 2 (**Figura 13**).





**Figura 12: Análisis estadístico de las diferencias en las proporciones relativas estimadas por MIXTURE entre los *clusters* definidos por agrupamiento no supervisado. Se informa el p-valor asociado a los contrastes entre *clusters*. Los resultados completos se muestran en la **tabla S.1** del Anexo.**



**Figura 13: Distribución de pacientes en los clusters obtenidos a partir del agrupamiento no supervisado y según su categoría de EMR al día 15.** Los clusters fueron definidos a partir de las proporciones estimadas por MIXTURE (LM22).

A continuación, se compararon las proporciones de las distintas poblaciones inmunitarias entre grupos de pacientes con distintas características clínicas. Cada población fue evaluada independientemente utilizando las proporciones estimadas a partir de TPM, ajustando a una distribución Beta, y utilizando como variable aleatoria el lote de secuenciación.

En la **tabla 6** se informan sólo las combinaciones de poblaciones celulares y variables clínicas en las que se detectaron diferencias estadísticamente significativas ( $p < 0,05$ ) entre los grupos de pacientes (los resultados completos se muestran en la **tabla S.3** del Anexo).

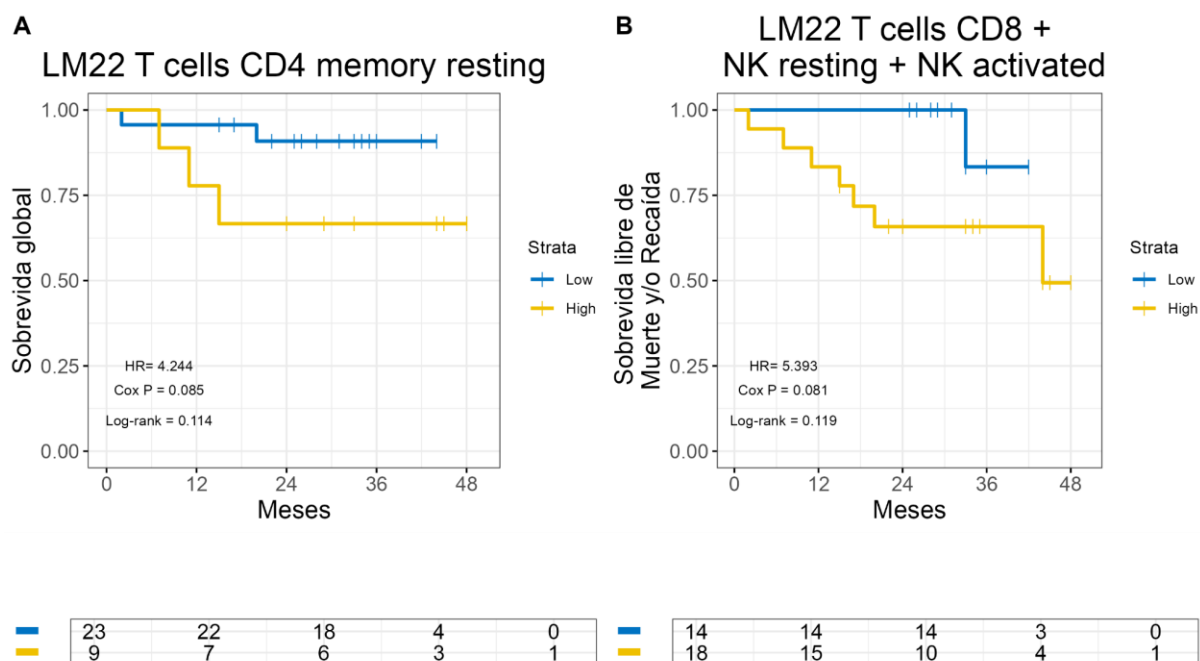
Población celular	Variable clínica	Media de proporción	p-valor
<i>T cells CD4<sup>+</sup></i> (TIL10)	Muerte y/o recaída	No: 0,0334 Si: 0,0775	0,0405
<i>Monocytes</i> (TIL10)	Muerte y/o recaída	No: 0,0244 Si: 0,0141	0,0234
<i>T cells CD8<sup>+</sup></i> (TIL10)	Toxicidad aguda relacionada al tratamiento	No: 0,0083 Si: 0,0169	$1,00 \times 10^{-6}$
<i>T cells CD8<sup>+</sup></i> (LM22)	Toxicidad aguda relacionada al tratamiento	No: 0,0048 Si: 0,0097	0,0438
<i>NK resting</i> (LM22)	Toxicidad aguda relacionada al tratamiento	No: 0,0067 Si: 0,0122	0,0085

**Tabla 6: Comparación de la abundancia de las poblaciones inmunitarias individuales entre grupos de pacientes con distintas características clínicas.** Sólo se incluyen aquellas diferencias estadísticamente significativas ( $p$ -valor  $< 0,05$ , distribución Beta con inflados en cero, variable aleatoria: lote de secuenciación).

Este análisis evidenció que los pacientes que presentaron toxicidad aguda relacionada al tratamiento presentaban una mayor proporción de linfocitos T CD8<sup>+</sup> (TIL10, LM22) y células

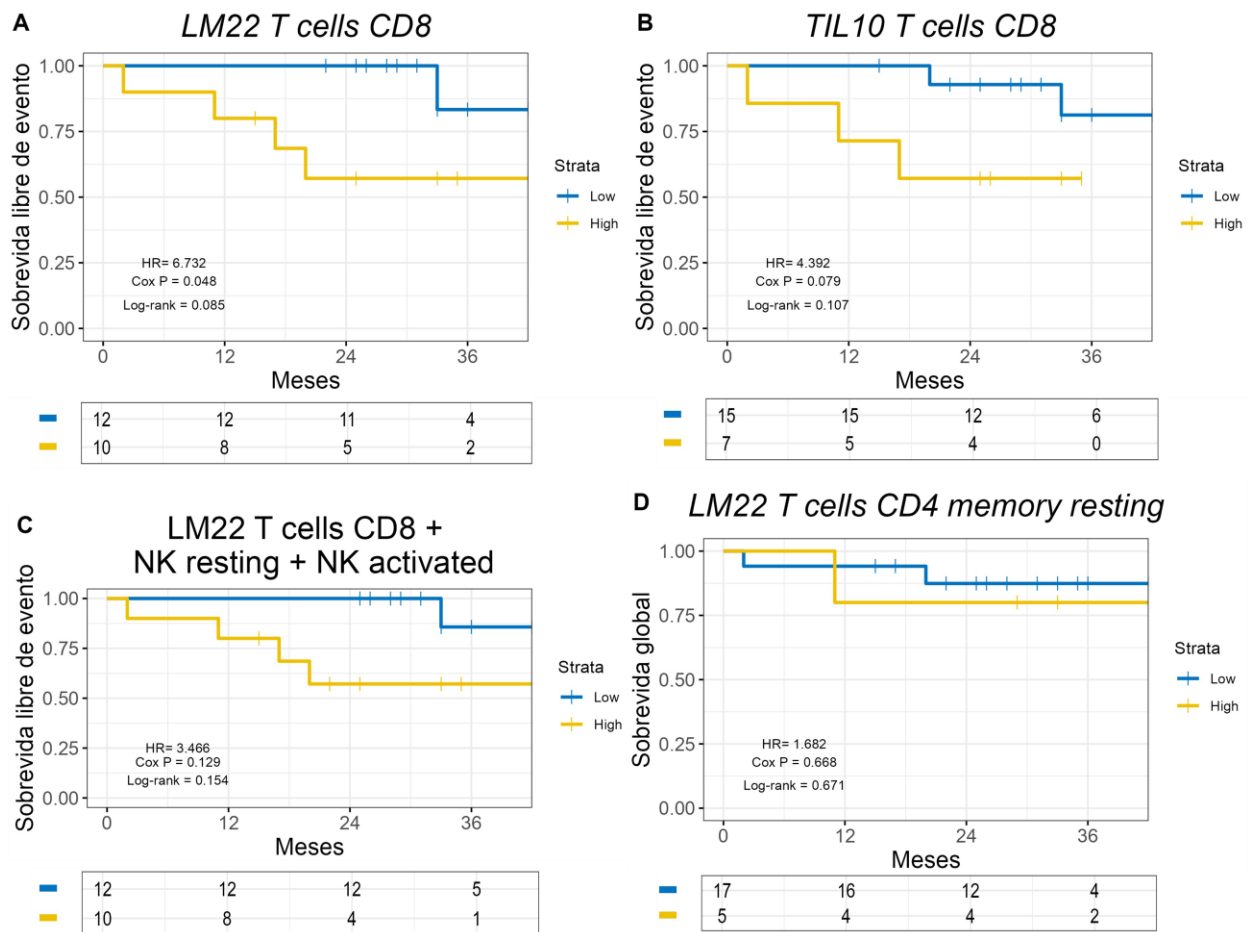
NK resting (LM22) (Tabla 6). En cuanto a muerte y/o recaída, encontramos que los pacientes que habían sufrido algún evento presentaban una mayor proporción de linfocitos T CD4<sup>+</sup> (TIL10) y una menor proporción de monocitos (TIL10) (Tabla 6). No se detectaron diferencias estadísticamente significativas para otras poblaciones celulares ni respecto a las variables clínicas restantes (tabla S.3).

Con el fin de evaluar si existe una asociación entre el componente inmunitario del TME y la sobrevida libre de evento (definida como el tiempo a la recaída o muerte, lo primero que ocurra), se realizaron curvas de Kaplan Meier comparando los pacientes con proporción alta vs. baja de cada población celular. El punto de corte fue determinado mediante la herramienta *Cutoff Finder* (Budczies, Klauschen, Sinn, Gyorffy, et al., 2012). En la tabla S.4 se informan el Hazard Ratio (HR), el p-valor de la prueba de Log-rank y el p-valor del HR (CoxP) para todas las curvas de sobrevida realizadas. El HR es una medida que representa la probabilidad que un individuo experimente un evento (ej. muerte) en un punto determinado de tiempo, de forma comparativa entre dos grupos. Un HR igual a 1 indica que no hay diferencia en el riesgo entre los dos grupos, mientras que un HR mayor a 1 indica un mayor riesgo en el grupo de interés en comparación con el grupo de referencia, y un HR menor a 1 indica un menor riesgo. Las poblaciones celulares están ordenadas de mayor a menor de acuerdo a su HR utilizando como evento Muerte (Tabla S.4). En la figura 14 se muestran las curvas de sobrevida que mostraron mayores diferencias entre grupos.



**Figura 14: Curvas de Kaplan Meier para la sobrevida de los pacientes según la proporción de las poblaciones celulares indicadas en el TME al momento del diagnóstico.** El valor de corte Low/High fue estimado mediante la herramienta *Cutoff Finder*. Los números en la parte inferior indican el número de pacientes en riesgo en cada punto de tiempo, “Low”: baja proporción (curvas azules); “High”: alta proporción (curvas amarillas).

En rasgos generales, se observó que los pacientes con mayores proporciones de linfocitos T CD4<sup>+</sup> o CD8<sup>+</sup> tenían mayor riesgo de presentar un evento. Esto sugiere que una mayor abundancia de células citotóxicas está asociada a una peor sobrevida en estos pacientes. Si bien las diferencias no fueron estadísticamente significativas, también hay que destacar que la cohorte de pacientes es pequeña y el número de eventos es bajo. Teniendo en cuenta que el grupo de riesgo tiene influencia en la probabilidad de recaída en esta patología, se repitió el análisis para cada grupo de riesgo por separado. Dado que los grupos de riesgo estándar y alto contenían un bajo número de pacientes, finalmente sólo se realizó para los pacientes de grupo intermedio (n=22). Los resultados que presentaron mayor HR se muestran en la **figura 15**.

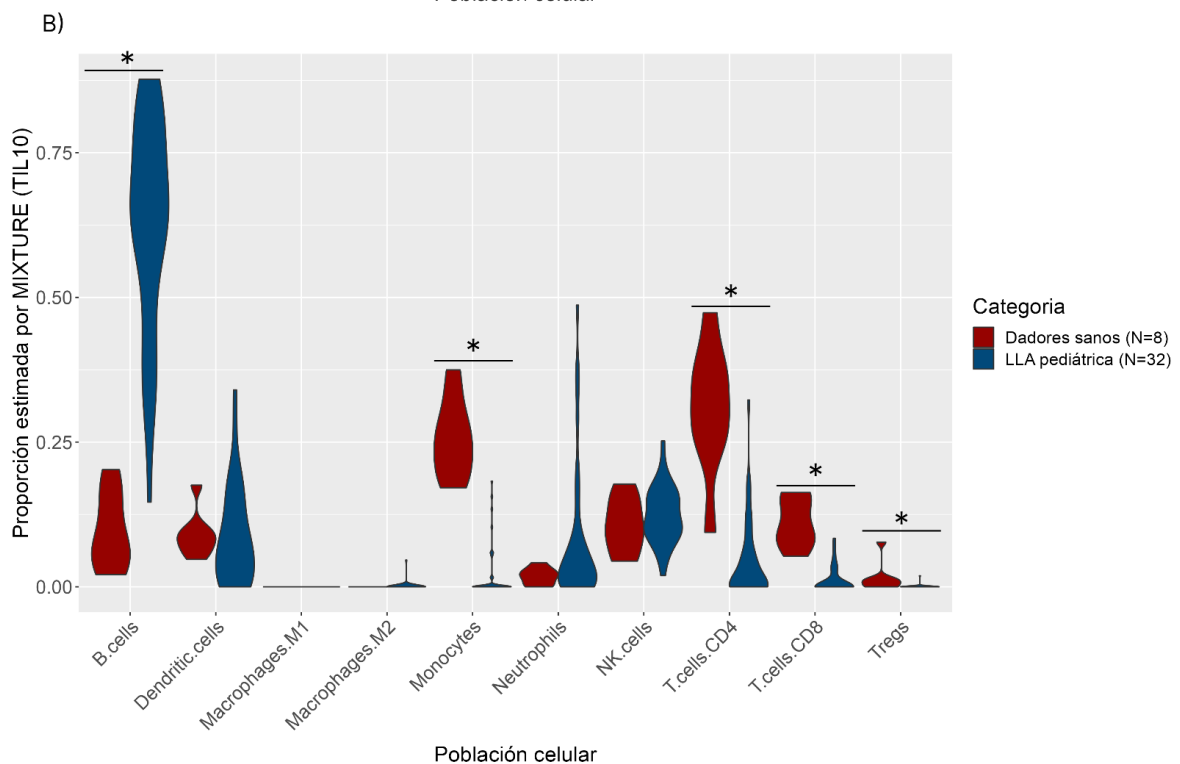
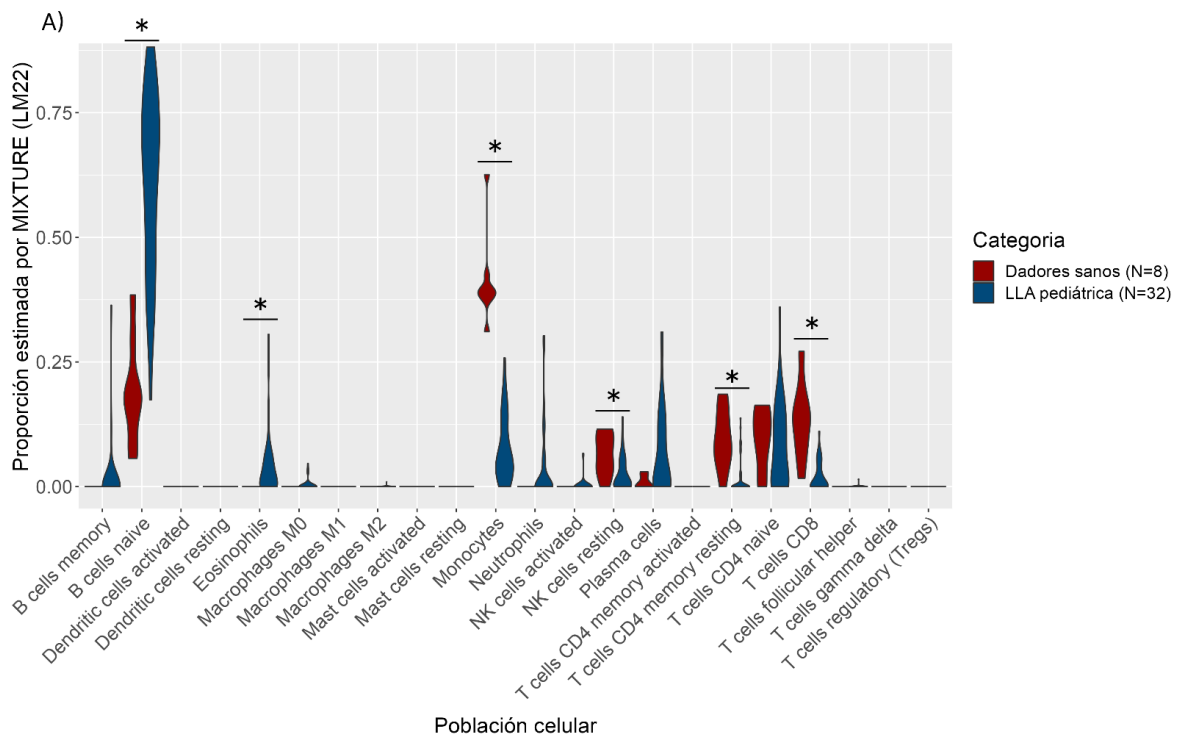


**Figura 15: Curvas de Kaplan Meier para la sobrevida de los pacientes pertenecientes al grupo de riesgo intermedio según la proporción de las poblaciones celulares indicadas en el TME al momento del diagnóstico.** El valor de corte Low/High fue estimado mediante la herramienta *Cutoff Finder*. Los números en la parte inferior indican el número de pacientes en riesgo en cada punto de tiempo, “Low”: baja proporción (curvas azules); “High”: alta proporción (curvas amarillas).

Se observó que, para pacientes clasificados como grupo de riesgo intermedio, una mayor proporción de linfocitos T CD8<sup>+</sup> está asociada a una menor sobrevida libre de muerte y/o recaída, ya sea que hayan sido estimadas a partir de la firma LM22 (HR=6,732; CoxP=0,048) o TIL10 (HR=4,392; CoxP=0,079). Esta tendencia se mantiene si se adicionan las células NK (HR=3,466; CoxP=0,129; **figura 15 C**). En cambio, las diferencias observadas previamente al separar a los pacientes según la abundancia de células T CD4<sup>+</sup> *memory resting* no se

mantuvieron al analizar sólo los pacientes con grupo de riesgo intermedio (HR=1,125; CoxP=0,899). Para el resto de poblaciones celulares no se detectaron asociaciones significativas con las variables clínicas analizadas.

Por último, se compararon las proporciones estimadas por MIXTURE de los pacientes de LLA con datos transcriptómicos provenientes de médula ósea de donantes sanos. Los datos correspondían a 8 donantes adultos, y se obtuvieron a partir del dataset GSE120444 (disponible en *Gene Expression Omnibus*). Cabe destacar que los datos provenían de la fracción de células mononucleares de médula ósea aisladas por gradiente de Ficoll, con lo cual no se deberían poder detectar granulocitos como neutrófilos, basófilos y eosinófilos. Como era de esperar, la proporción de células B (TIL10) o B *naive* (LM22) era mayor en pacientes de LLA, dado que en éstos dichas poblaciones representan en su mayoría a los blastos leucémicos (**figura 16, tabla S.5**). Otras poblaciones en las que se diferenciaban las proporciones fueron los monocitos (ambas matrices), linfocitos T CD4<sup>+</sup> (TIL10), linfocitos T CD4<sup>+</sup> *memory resting* (LM22) y linfocitos T CD8<sup>+</sup> (ambas matrices), que resultaron más abundantes en donantes sanos (prueba de Wilcoxon, p<0,05). Dado que se trata de proporciones relativas, es difícil saber qué tanto de estas diferencias se deben a la abundancia real de estas poblaciones y no a la carencia de granulocitos y de blastos leucémicos en las muestras de donantes sanos. (Blaeschke et al., 2020). Aun así, se puede resaltar que no se encontraron diferencias estadísticamente significativas en la proporción relativa de células NK (TIL10), células dendríticas (TIL10), células T CD4<sup>+</sup> *naive* (LM22) (prueba de Wilcoxon, p>0,05), e incluso una mayor proporción relativa de células NK *resting* (LM22, prueba de Wilcoxon, p=0,019) en pacientes con LLA, sugiriendo que el número absoluto de dichas poblaciones en pacientes es mayor que en médula sana, por el efecto de las células de linaje B sobre las proporciones relativas de otras poblaciones celulares en los pacientes.

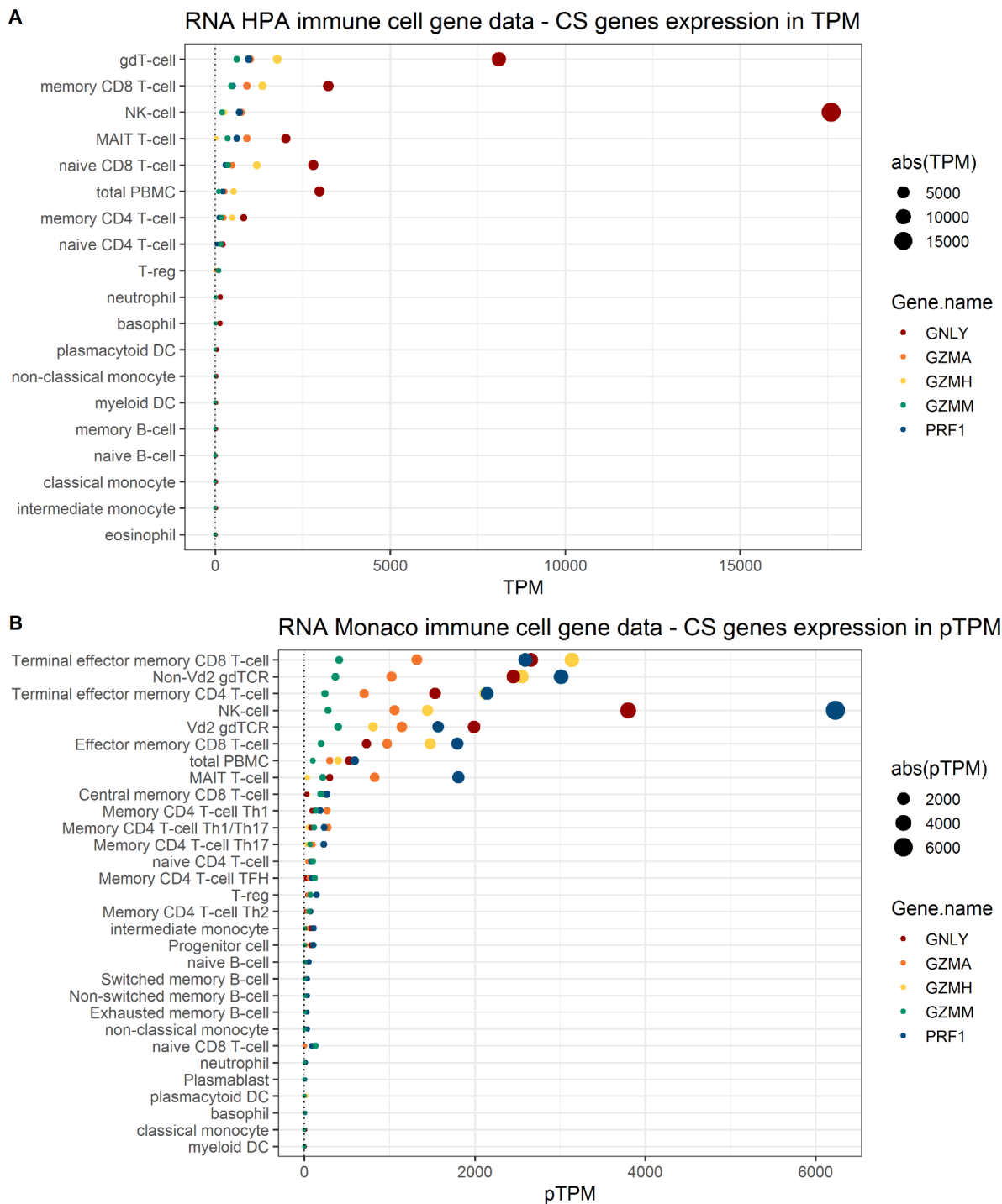


**Figura 16: Médula ósea sana vs. LLA-B.** Comparación entre las proporciones celulares del componente inmunitario en médula ósea estimados por MIXTURE, a partir de pacientes con LLA-B (n=32) vs. donantes adultos sanos (n=8). \*: p-valor < 0,05. Los resultados de la prueba de Wilcoxon se muestran en la **tabla S.5** del Anexo.

### IV.3) *Cytolytic Score*: cálculo y validación

Con el fin de complementar la caracterización del microambiente tumoral de los pacientes, implementamos un estimador de la abundancia de células citotóxicas (células NK y linfocitos T citotóxicos), al cual llamaremos *cytolytic score* (CS), a partir de los niveles de expresión de 5 genes con relevancia funcional en este tipo de células: *GNLY*, *GZMA*, *GZMH*, *GZMM* y *PRF1*. Este índice fue propuesto como una modificación de versiones anteriores del CS en un reciente trabajo (Dufva et al., 2020), realizado exclusivamente con muestras de neoplasias hematológicas. Para validar esta estrategia, los autores analizaron la fracción de linfocitos T y NK medidos por citometría de flujo en muestras de médula ósea de pacientes con leucemia mieloide aguda y posteriormente realizaron RNA-seq. El CS se correlacionó positivamente con la abundancia combinada de linfocitos T y NK de entre todas las células de la médula ósea. Adicionalmente, validaron el CS mediante la cuantificación de células T por inmunohistoquímica en muestras de linfoma difuso de células grandes B y linfoma de tejido linfoide asociado a mucosas. Concluyeron que el CS estimaba robustamente la abundancia de células T citotóxicas y NK en neoplasias hematológicas (Dufva et al., 2020).

En esta tesis nos propusimos confirmar la especificidad del CS, por lo que analizamos el nivel de expresión de los 5 genes del CS utilizando los datos disponibles en el repositorio público *The Human Protein Atlas* (<https://www.proteinatlas.org/>). En la **figura 17**, se muestra que los 5 genes incluidos en CS se expresan en altos niveles principalmente en linfocitos T (CD8<sup>+</sup>, CD4<sup>+</sup> y gamma-delta) y células NK. Entre los linfocitos T se detectan particularmente en un perfil de memoria efector. Es de interés observar que estos genes también se expresan en al menos un subtipo de linfocitos T CD4<sup>+</sup> los cuales, tradicionalmente, no se los asocia con un fenotipo citotóxico. Por otro lado, según estos datos obtenidos de *The Human Protein Atlas*, los genes de mayor expresión son *GNLY* y *PRF1*, mientras que *GZMA* y *GZMH* tienen una expresión intermedia y *GZMM*, en general, se expresa en niveles mucho más bajos que el resto de los genes.

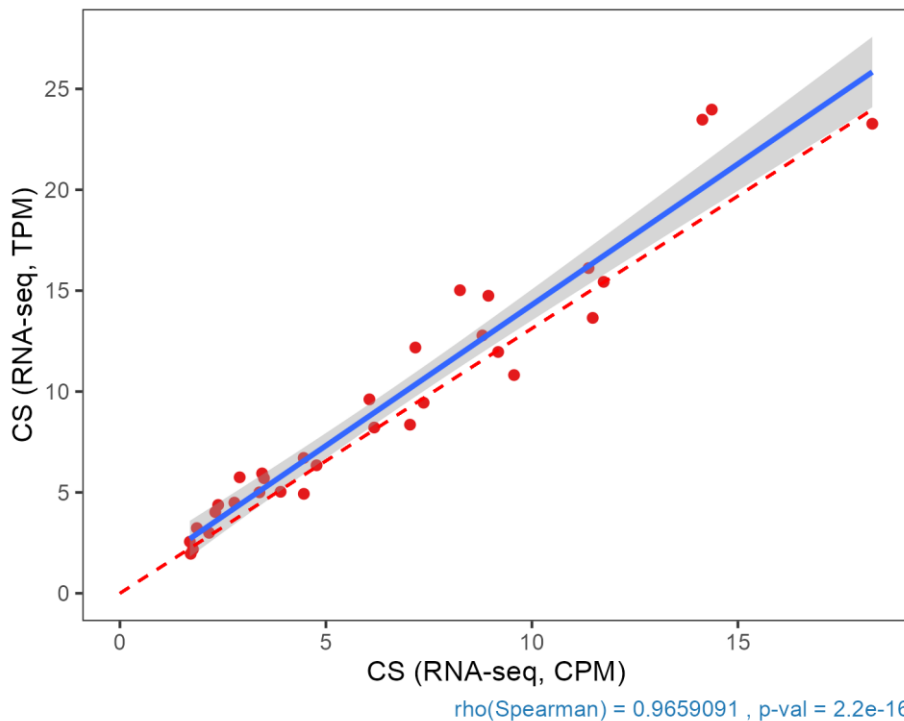


**Figura 17: Nivel de expresión de los 5 genes del Cytolytic Score (*GNLY*, *GZMA*, *GZMH*, *GZMM* y *PRF1*) en distintas poblaciones inmunitarias según 2 bases de datos de The Human Protein Atlas. A: base de datos “RNA HPA immune cell gene data”. B: base de datos “RNA Monaco immune cell gene data”. gdT cells: células T gamma-delta; MAIT T-cells: Mucosal associated invariant T-cells; PBMC: Peripheral blood mononuclear cells; DC: dendritic cell. pTPM: protein coding transcripts per million.**



### IV.3.1) Caracterización del CS en muestras de LLA-B a partir de datos transcriptómicos (RNA-seq)

Se calculó el CS en las muestras de los pacientes argentinos a partir de los datos de TPM o CPM de RNA-seq, y se compararon ambas métricas con el objetivo de continuar con el análisis con sólo una de ellas. Los valores de CS presentaron una correlación significativa ( $\rho$  Spearman = 0,965,  $p$ -valor =  $2,2 \times 10^{-16}$ ) (**figura 18**), y una dispersión muy solapada con la línea de identidad, sugiriendo que sería equivalente usar cualquiera de estas dos métricas.



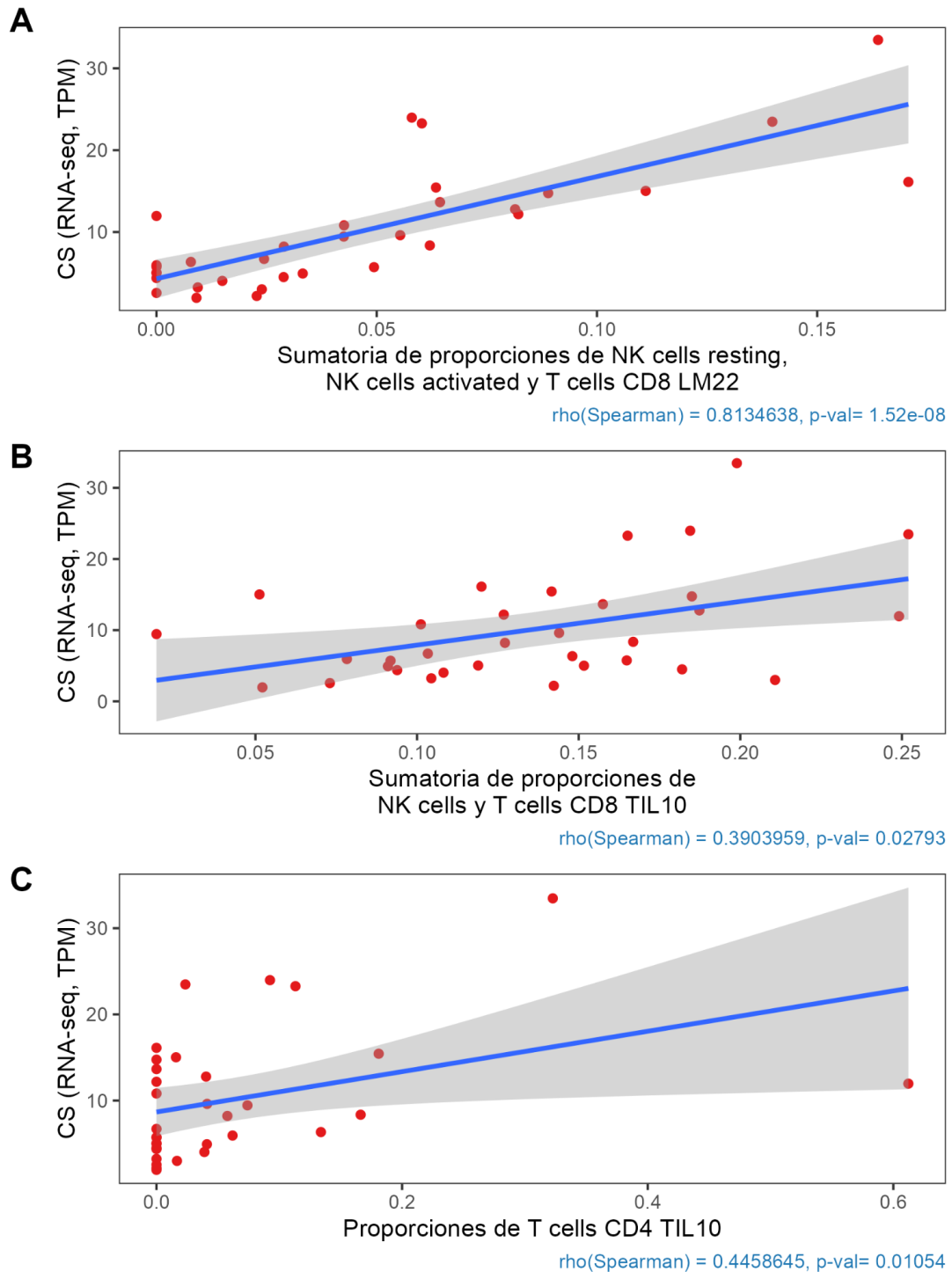
**Figura 18: Comparación del CS según las unidades de cuantificación de transcritos.** Gráficos de correlación entre CS calculado a partir de los niveles de expresión de los genes *GNLY*, *PRF1*, *GZMH*, *GZMA* y *GZMM* en muestras de LLA-B, determinados por RNA-seq y cuantificados en unidades de TPM (eje y) o CPM (eje x).

Se comparó el CS con respecto a las proporciones celulares estimadas por MIXTURE, tanto por LM22 como por TIL10. En términos generales, se observó una correlación positiva significativa entre el CS y células NK, linfocitos T CD8<sup>+</sup> y linfocitos T CD4<sup>+</sup> (**Tabla 7**). La sumatoria de los linfocitos T CD8<sup>+</sup> y las células NK resultó en mayores coeficientes de correlación con el CS (LM22:  $\rho$  Spearman = 0,813,  $p$ -valor < 0,001; TIL10:  $\rho$  Spearman = 0,390,  $p$ -valor = 0,028) (**figura 19**), aunque consideramos importante tener en cuenta que existe un aporte dado por las células T CD4<sup>+</sup>. Dado que la matriz LM22 incluye los genes del CS mientras que TIL10 no, sólo la comparación con esta última es independiente. Los resultados obtenidos sugieren que el CS puede utilizarse como estimador de la abundancia de células citotóxicas en muestras de LLA-B.

**Tabla 7: Resultados de correlación entre CS (RNA-seq, TPM) y proporciones relativas de poblaciones celulares estimadas por MIXTURE.**

<b>CS RNA-seq TPM</b>		
<b>Población inmunitaria TIL10</b>	<b>Rho (Spearman)</b>	<b>p-valor</b>
Linfocitos T CD4 <sup>+</sup>	0,4458645	0,01054
Linfocitos T CD8 <sup>+</sup>	0,3947373	0,02536
Sumatoria Linfocitos T CD8 <sup>+</sup> y NK	0,3903959	0,02793
Células B	-0,2998534	0,09566
Células NK	0,2397361	0,1857
Neutrófilos	0,07895225	0,6675
Células Dendríticas	-0,06674669	0,7166
Monocitos	0,04819748	0,7934
<b>CS RNA-seq TPM</b>		
<b>Población inmunitaria LM22</b>	<b>Rho (Spearman)</b>	<b>p-valor</b>
Sumatoria Linfocitos T CD8 <sup>+</sup> y NK	0,8134638	1,52x10 <sup>-8</sup>
Células NK <i>resting</i>	0,4695848	0,006697
Linfocitos B <i>naive</i>	-0,4380499	0,01283
Linfocitos T CD8 <sup>+</sup>	0,4149126	0,01821
Neutrófilos	0,3005113	0,09468
Células Plasmáticas	0,2542333	0,1603
Eosinófilos	-0,1156302	0,5286
Linfocitos T CD4 <sup>+</sup> <i>memory resting</i>	-0,1109253	0,5456
Macrófagos M0	0,1014174	0,5807
Monocitos	-0,02880207	0,8757
Linfocitos T CD4 <sup>+</sup> <i>naive</i>	0,02494013	0,8922

**Tabla 7:** Correlación entre poblaciones celulares estimadas por MIXTURE (TIL10 y LM22) y el CS calculado con los TPM del RNA-seq. Se presentan el Rho de Spearman y el p-valor. Solo se incluyen las poblaciones celulares para las que al menos 3 muestras tenían proporciones distintas de cero.



**Figura 19: Análisis de correlación entre el CS y poblaciones estimadas por MIXTURE.** Gráficos de correlación entre el CS calculado a partir del RNA-seq (TPM) y las proporciones estimadas de células T CD8<sup>+</sup>/NK por MIXTURE (A: LM22, B: TIL10) y células T CD4<sup>+</sup> (C: TIL10).

### IV.3.2) Ensayo de cuantificación del CS mediante RT-qPCR

Dada la utilidad de estimar la abundancia de células citotóxicas con los niveles de RNA de sólo 5 genes, decidimos poner a punto una metodología que permita la cuantificación del CS sin requerir la secuenciación del transcriptoma. Para ello se diseñaron *primers* específicos para cada gen (**Figura S.2**) y se puso a punto su detección utilizando la técnica de RT-qPCR

utilizando RNA de células mononucleares activadas de sangre periférica de donantes sanos (DS661 y DS662).

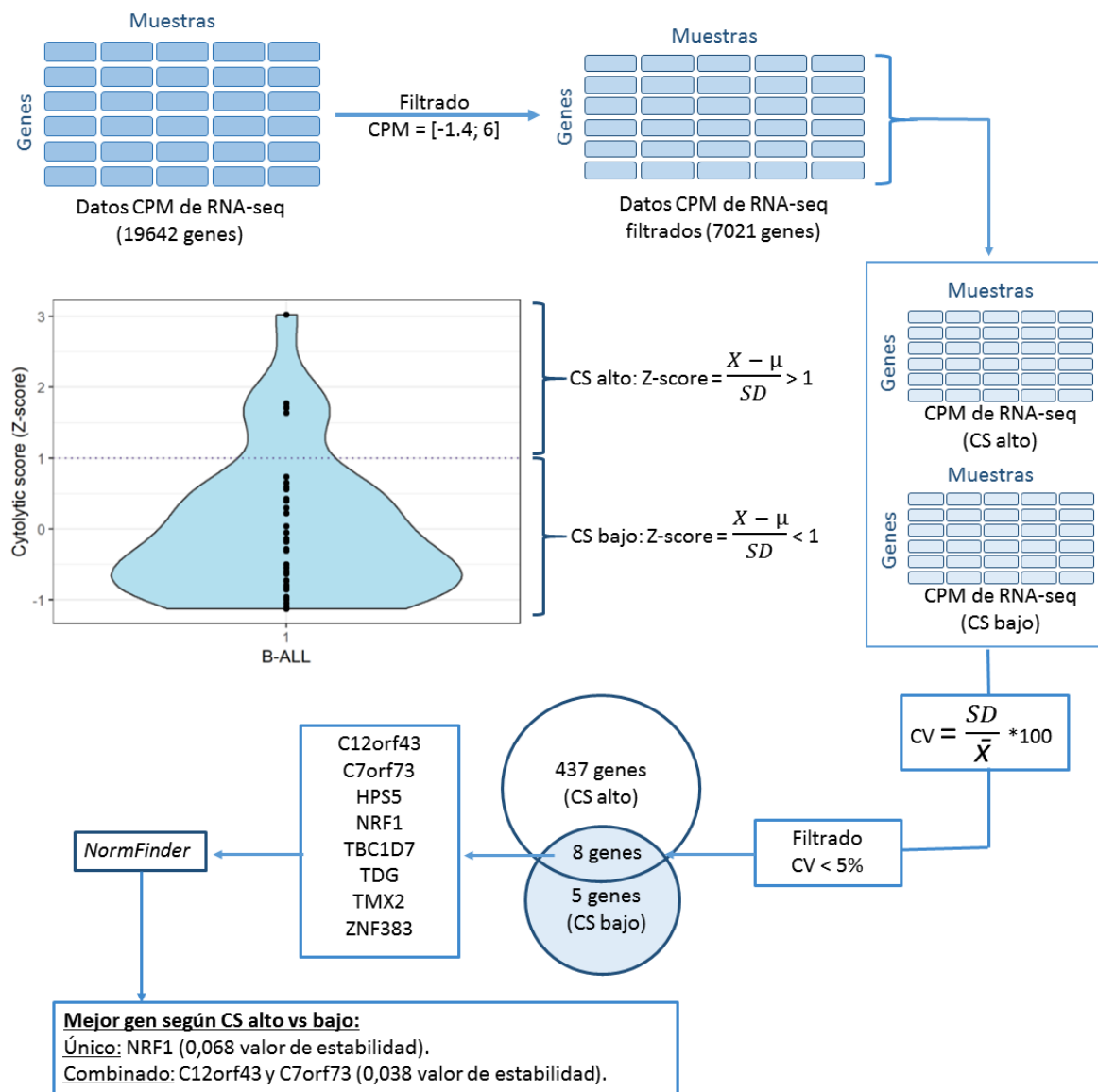
Con los *primers* de *GNLY*, *GZMA* y *PRF1* se obtuvieron bandas intensas del tamaño esperado en todas las temperaturas testeadas. *GZMH* presentó bandas muy tenues en todas las temperaturas excepto 57°C, y *GZMM* no presentó banda en ninguna condición. Como *GNLY*, *PRF1*, *GZMA* y *GZMH* presentaron sus respectivas bandas (aunque *GZMH* muy tenue) del tamaño esperado (ver Materiales y Métodos) a una temperatura de annealing de 60°C, se decidió proseguir con los experimentos utilizando esta temperatura de *annealing*.

Para los *primers* de *GZMM* y *GZMH* se probaron distintas concentraciones de MgCl<sub>2</sub> (2 y 2,5 mM) y de DMSO (2 y 5% del volumen total) en la Mix de PCR punto final para mejorar la amplificación del segmento de cDNA deseado. *GZMH* presentó banda intensa utilizando DMSO 2% y MgCl<sub>2</sub> 2 mM, mientras que *GZMM* no amplificó un producto del tamaño esperado en ninguna condición. En ambos casos (*GZMH* y *GZMM*) se observaron dímeros de *primers*.

Dado que *GZMM* no amplificó la secuencia deseada de cDNA del gen en cuestión decidimos proseguir con el cálculo del CS sin utilizar este gen. Con lo cual los genes que compondrían al CS serían *GNLY*, *GZMA*, *GZMH* y *PRF1*.

### IV.3.3) Búsqueda de genes de Referencia

Paralelamente a esto, se seleccionaron genes de referencia que pudiesen utilizarse para calcular la expresión relativa de los genes del CS al momento de realizar la RT-qPCR. Para ello primeramente se realizó un z-score para dividir a los pacientes que tenían un CS alto y CS bajo, quedando 4 pacientes en el grupo de CS alto (z-score > 1). Luego se calculó un coeficiente de variación (CV) para los genes de expresión similar a los 5 genes del CS para cada grupo por separado y se retuvieron los genes que tenían un CV < 5%. Se obtuvo un listado de 13 genes en el grupo de CS bajo y 445 en el grupo de CS alto; siendo 8 los genes en común: *C12orf43*, *C7orf73*, *HPS5*, *NRF1*, *TBC1D7*, *TDG*, *TMX2* y *ZNF383*. A continuación, se utilizó la herramienta *NormFinder* (Andersen et al., 2004) para encontrar entre esos 8 genes cuáles eran los óptimos. El procedimiento de filtrado de genes se resume en la **figura 20**.



**Figura 20: Diagrama de flujo de la selección de genes de referencia hasta utilizar el programa de NormFinder.** CV = coeficiente de variación; SD = desvío standard;  $\bar{x}$  = media aritmética; CS = cytolitic score; X = score;  $\mu$  = media.

El programa permite seleccionar un “grupo identificador”, se probaron los grupos “CS alto y bajo”, “Recaída”, “Muerte” y “Toxicidad aguda relacionada al tratamiento”. Los resultados se muestran en la **Tabla 8:**

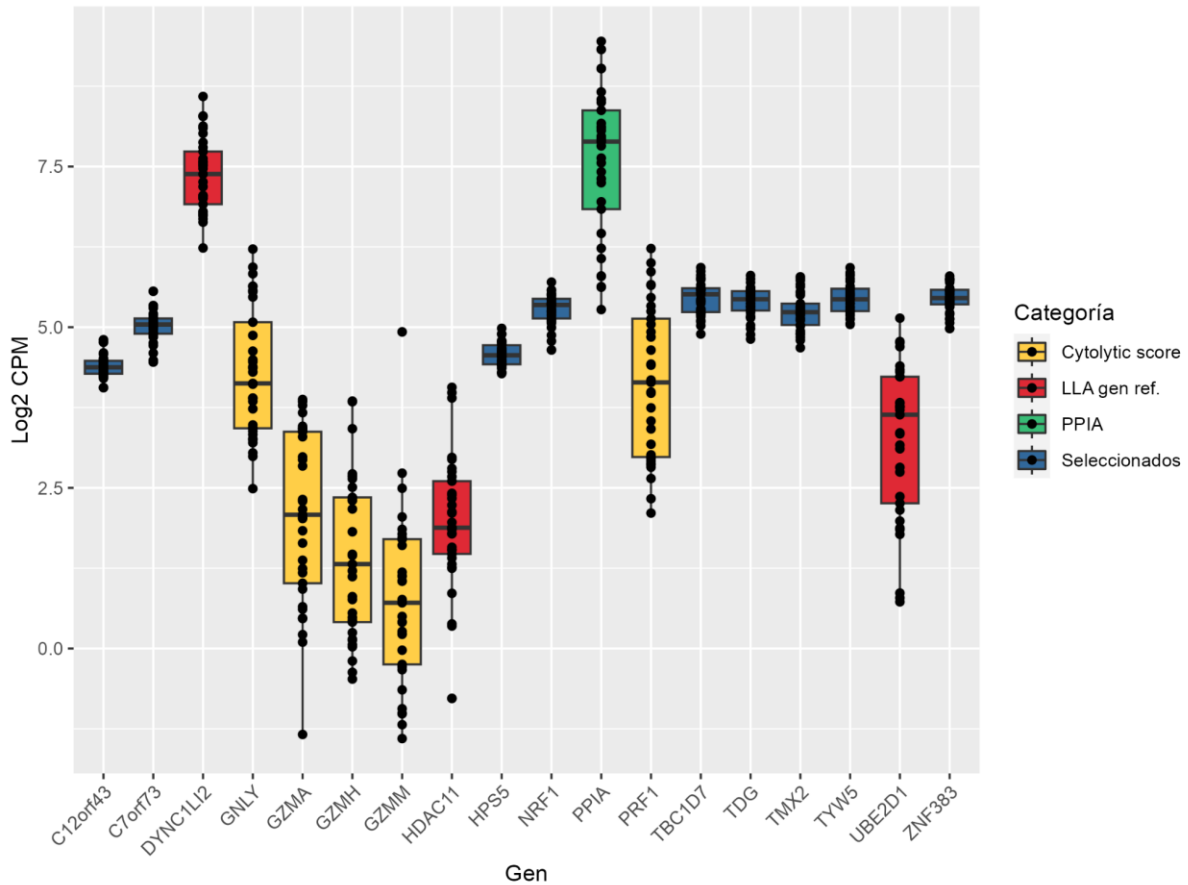
**Tabla 8: Resultados de NormFinder para la búsqueda del gen de referencia.**

Grupo identificador: CS alto vs bajo	
Mejor gen	C12orf43
Valor de estabilidad	0,057

Mejor combinación de 2 genes	C12orf43 y NRF1
Valor de estabilidad para mejor combinación de 2 genes	0,046
<b>Grupo identificador: Recaída</b>	
Mejor gen	NRF1
Valor de estabilidad	0,068
Mejor combinación de 2 genes	C12orf43 y C7orf73
Valor de estabilidad para mejor combinación de 2 genes	0,038
<b>Grupo identificador: Muerte</b>	
Mejor gen	C12orf43
Valor de estabilidad	0,070
Mejor combinación de 2 genes	C12orf43 y C7orf73
Valor de estabilidad para mejor combinación de 2 genes	0,038
<b>Grupo identificador: Toxicidad aguda asociada a tratamiento</b>	
Mejor gen	C12orf43
Valor de estabilidad	0,039
Mejor combinación de 2 genes	C12orf43 y HPS5
Valor de estabilidad para mejor combinación de 2 genes	0,032

**Tabla 8:** Resultados de NormFinder para la búsqueda del gen de referencia. Para cada grupo identificador se indica el mejor gen individual y combinación de 2 genes junto con sus valores de estabilidad.

Para visualizar mejor los resultados se graficó la expresión de los genes de referencia seleccionados junto con los 5 genes del CS y otros genes de referencia cuyos *primers* ya se habían diseñado en el laboratorio y podíamos contar con ellos en caso de que tengan una expresión similar a la de los 5 genes del CS para todas las muestras (**figura 21**).



**Figura 21: Comparación de los niveles de expresión de genes de referencia candidatos en muestras de LLA-B, medidos por RNA-seq.** Expresión de genes del CS (amarillo); genes de referencia de los cuales ya se tenían primers diseñados en el laboratorio (rojo); PPIA (verde), genes filtrados por medio de su CV y que eran comunes en los pacientes de CS alto y bajo (azul).

En función de los resultados de *NormFinder* y el patrón de expresión observado en las muestras de LLA, seleccionamos los genes *NRF1*, *HPS5* y *DYNC1LI2* como candidatos a genes de referencia. También evaluamos el desempeño de *PPIA* por ser ya utilizado en el laboratorio para el estudio de líneas celulares. Descartamos en esta primera instancia a *C12orf43* por tratarse de un gen con función desconocida, y porque encontramos muchas dificultades para diseñar *primers* específicos.

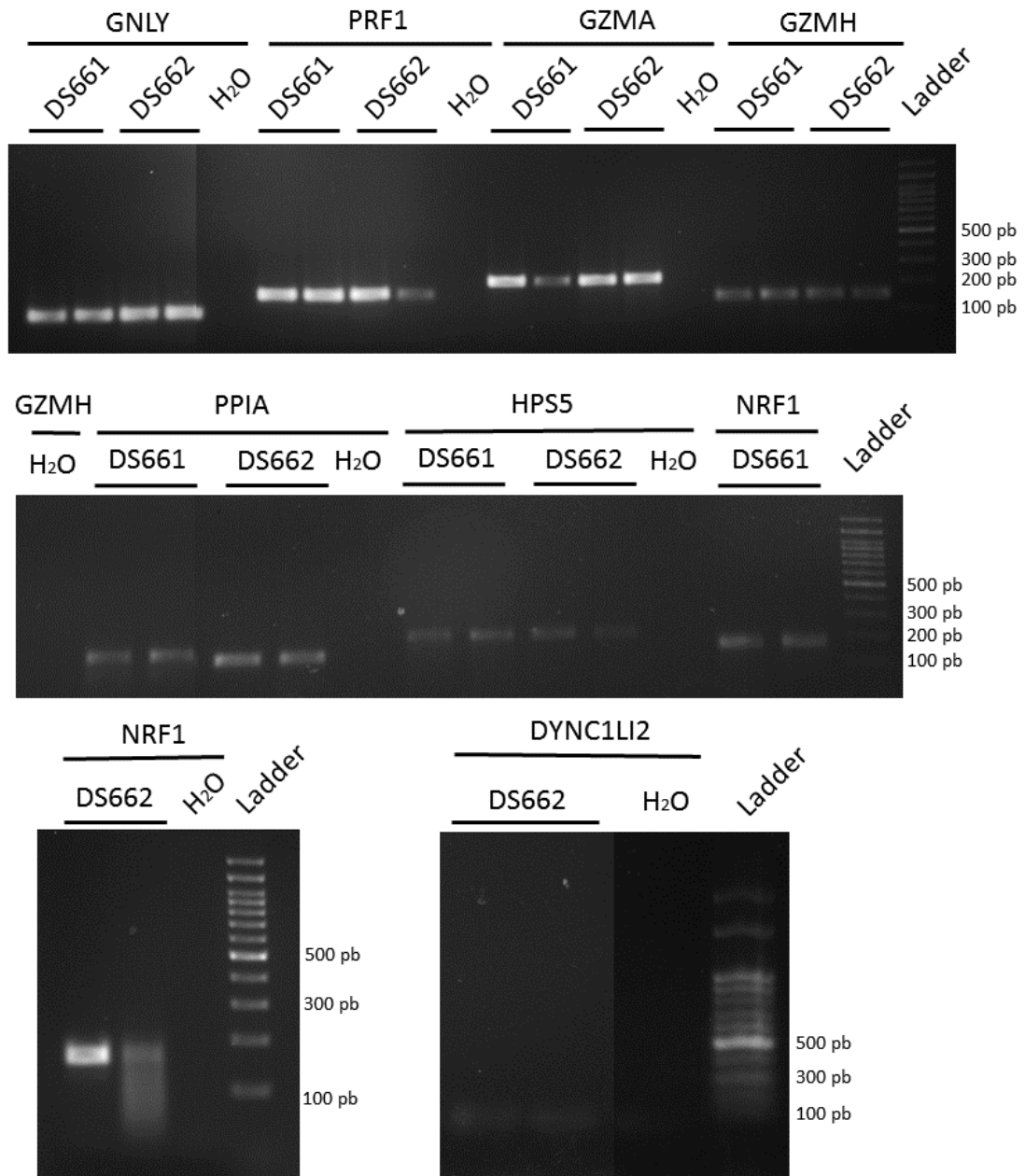
A partir de los resultados obtenidos mediante el análisis bioinformático y de la puesta a punto de la RT-qPCR, se comparó el desempeño de los genes de referencia seleccionados (**Tabla 9, figura 22**).

**Tabla 9: Consideraciones para la selección del gen de referencia.**

	<i>HPS5</i>	<i>DYNC1LI2</i>	<i>NRF1</i>	<i>PPIA</i>
Poca variación de expresión entre muestras (RNA-seq)	Si	Si	Si	No
Expresión similar a la de los genes del CS (RNA-seq)	Si	No	Si	No

Un solo pico en la curva de melting (RT-qPCR)	No	No	Si	Si
Amplifica producto específico (RT-PCR y RT-qPCR)	Si	Sí	Si	Si

**Tabla 9:** Consideraciones para la selección del gen de referencia.

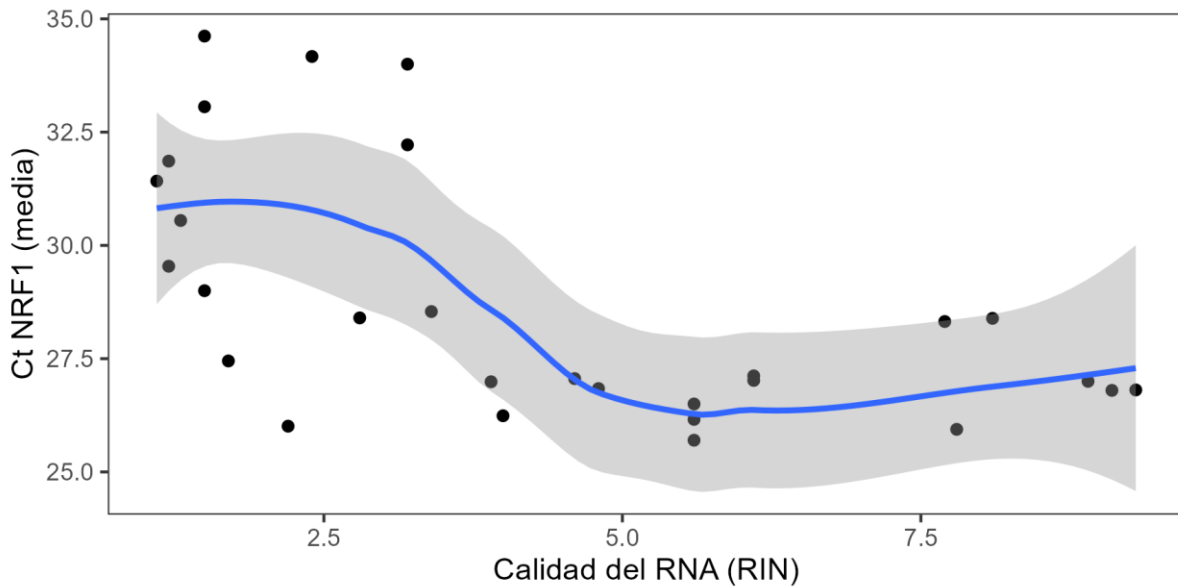


**Figura 22: Puesta a punto de la detección de los genes del CS y genes de referencia por RT-qPCR.** Electroforesis en gel de agarosa de productos de RT-qPCR de *GNLY*, *GZMA*, *GZMH*, *PRF1*, *NRF1*, *PPIA*, *DYNC1LI2* y *HPS5*, utilizando como cDNA molde las muestras de DS661, DS662 y H<sub>2</sub>O como control negativo, sembrados en gel de agarosa junto con *ladder* de 100 pb.



Si bien *HPS5* parecía un buen candidato, presentaba un doble pico en la curva de *melting*, con temperaturas muy cercanas, descartando que se tratase de dímeros de *primers*. Tras enviar a secuenciar el producto de PCR se confirmó que se trataba de un producto único. Concluimos que el doble pico podía deberse a una característica del amplicón, el cual presenta regiones con mucha diferencia en su temperatura de *melting*. Descartamos el gen de referencia *PPIA* ya que el Ct era muy elevado (Ct entre 29,50 y 30,36 en muestras de donantes sanos diluidas 1/20), no sólo en las muestras de donante sano, sino en otras muestras de LLA evaluadas previamente en el laboratorio. Planteamos que la discordancia entre los niveles de *PPIA* detectados por RNA-seq y RT-qPCR podrían deberse a que los *primers* de *PPIA* no amplifican todas las isoformas de RNAm. Por estas razones y debido a que no cumplía con el resto de características a considerar para usarlo como gen de referencia, se descartó como candidato. Por último, *DYNC1L2* no amplificó producto de PCR para las muestras de donantes sanos. Sin embargo, sí lo hizo para una línea celular de precursores de linfocitos B, TOM1, en donde se vió un doble pico en la curva de *melting* y al sembrar el producto de qPCR en un gel de agarosa se observó una banda de tamaño esperado. Por lo tanto, los *primers* diseñados sí eran específicos del gen deseado, pero fue descartado por no expresarse en donantes sanos.

Por último, se seleccionaron algunas muestras representativas de las 32 muestras totales en cuanto a su CS (calculado a partir de los datos del RNA-seq) para verificar la estabilidad de expresión de los posibles genes de referencia. Las muestras seleccionadas fueron POS002 y HLP013 (como representantes de CS alto), HNG008 y POS026 (como representantes de CS intermedio) y HNG003, HLP012 y POS025 (como representantes de CS bajo), todas por triplicado. Se pudo detectar expresión de todos los genes en todas las muestras a excepción de HLP012, pero con niveles variables de Ct. Dado que las muestras de pacientes, a pesar de tener una concentración de RNA similar al momento de la retrotranscripción, presentaban una calidad muy variable (**figura S.1**), decidimos no descartar ningún gen por este resultado. En particular para *NRFI* se evaluó la relación entre el Ct y la calidad de la muestra de RNA a través del parámetro RIN medido por *Bioanalyzer*, y se observó que aquellas muestras con RIN<4 presentaron mayor variabilidad y una dispersión alrededor de Ct~30, mientras que aquellas con RIN>4 presentaron menor variabilidad y una dispersión alrededor de Ct~26 (**Figura 23**). Este patrón permitió explicar parcialmente la variabilidad observada en *NRFI*, sin embargo, no podemos descartar otros factores presentes o variaciones menores en los niveles de expresión del gen.



**Figura 23: Relación entre Ct y calidad de la muestra de RNA para el gen de referencia *NRF1*.** Media de los valores de Ct medidos por RT-qPCR vs. el parámetro de calidad RIN medido por Bioanalyzer en el RNA, para las muestras de LLA-B incluidas en este trabajo. En azul se indica la línea de tendencia (método LOESS) y en gris el intervalo de confianza (95%).

En base a estos resultados y al análisis realizado por *NormFinder* decidimos utilizar *NRF1* como gen de referencia para el cálculo del CS por RT-qPCR.

Una vez seleccionado el gen de referencia y los genes para estimar el CS, se calculó la eficiencia de amplificación y el rango dinámico de cuantificación para todos ellos. Los resultados se informan en la **Tabla 10**:

**Tabla 10: Eficiencia de amplificación por RT-qPCR.**

Gen	Eficiencia (%)	Número de puntos de la curva	Rango dinámico (Ct)
<i>GNLY</i>	87,35	4	21,39 - 32,61
<i>GZMA</i>	69,57	3	24,36 - 33,08
<i>GZMH</i>	72,53	3	24,48 - 32,93
<i>PRF1</i>	107,79	3	23,22 - 29,52
<i>NRF1</i>	87,27	3	27,97 - 35,31

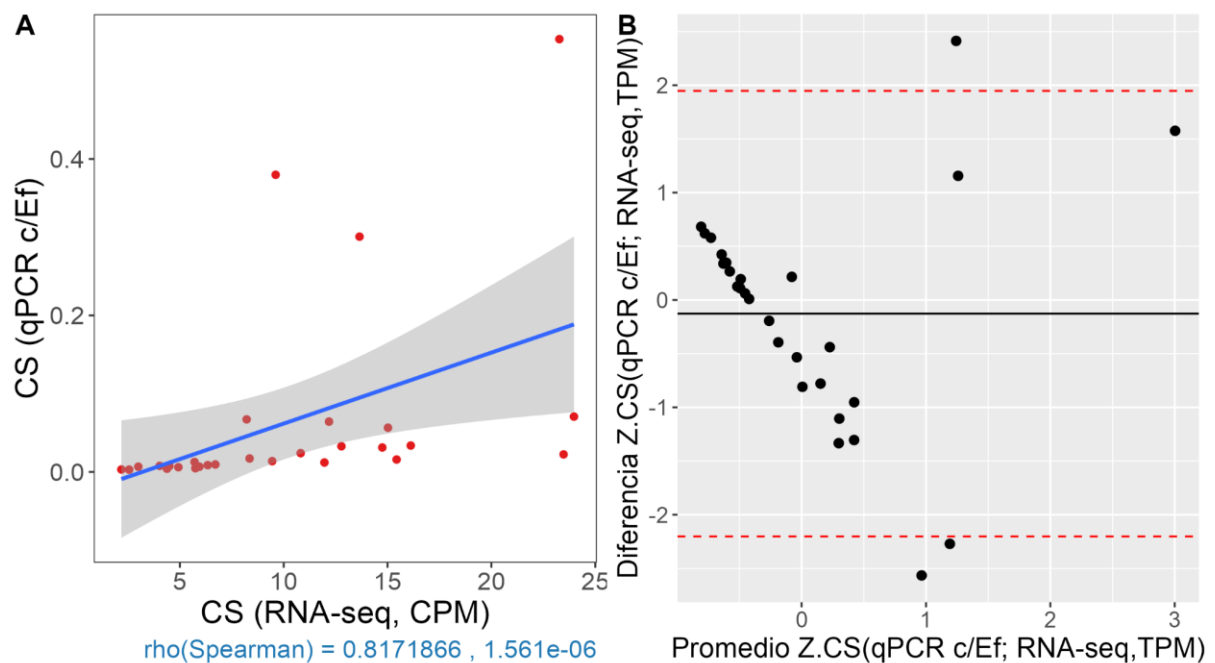
**Tabla 10:** Eficiencia de amplificación estimada según se especifica en Materiales y métodos. Las curvas con 3 puntos resultaron de la no amplificación en la mayor dilución de la curva.

Debido a que no se obtuvieron eficiencias de amplificación óptimas, se decidió calcular la expresión génica relativa teniendo en cuenta la eficiencia de amplificación para cada gen según se describe en materiales y métodos.

### IV.3.4) RT-qPCR de muestras de pacientes de LLA-B

Debido a que las muestras no podían analizarse todas en una única reacción de qPCR, se incluyó la muestra de referencia DS662 en todas las reacciones de manera de poder corregir los resultados por la variabilidad inter-reacción. Se excluyó a la muestra POS013 por ser insuficiente la cantidad de RNA para realizar las 5 reacciones requeridas para el cálculo del CS.

Se calculó el CS para cada muestra según se indica en Materiales y Métodos. Los valores de CS estimados por RT-qPCR presentaron una alta correlación positiva (Rho de Spearman = 0,8171866, p-valor =  $1,56 \times 10^{-6}$ ) (**figura 24**, izquierda) y concordancia (estimada por Bland Altman) con los datos de CS estimados por RNA-seq (**figura 24**, derecha). Sin embargo, el patrón de puntos observado en el análisis de Bland Altman, no simétricamente distribuido alrededor de la línea media, sugiere una estructura o sesgo de los datos, requiriendo mayor validación en el futuro. Por otro lado, el CS de las muestras HPL009, HPL011, POS027 y HPL012 no se pudo calcular ya que al menos un gen del CS no mostró expresión detectable. Se pudo calcular el CS en 28/32 muestras utilizando todos los genes (*GNLY*, *GZMA*, *GZMH* y *PRF1*).



**Figura 24: Comparación entre CS calculado a partir de RNA-seq y RT-qPCR.** Gráfico de correlación (A). Gráfico de concordancia según Bland Altman (B). Z.CS: Z-score del CS calculado por ambas metodologías.

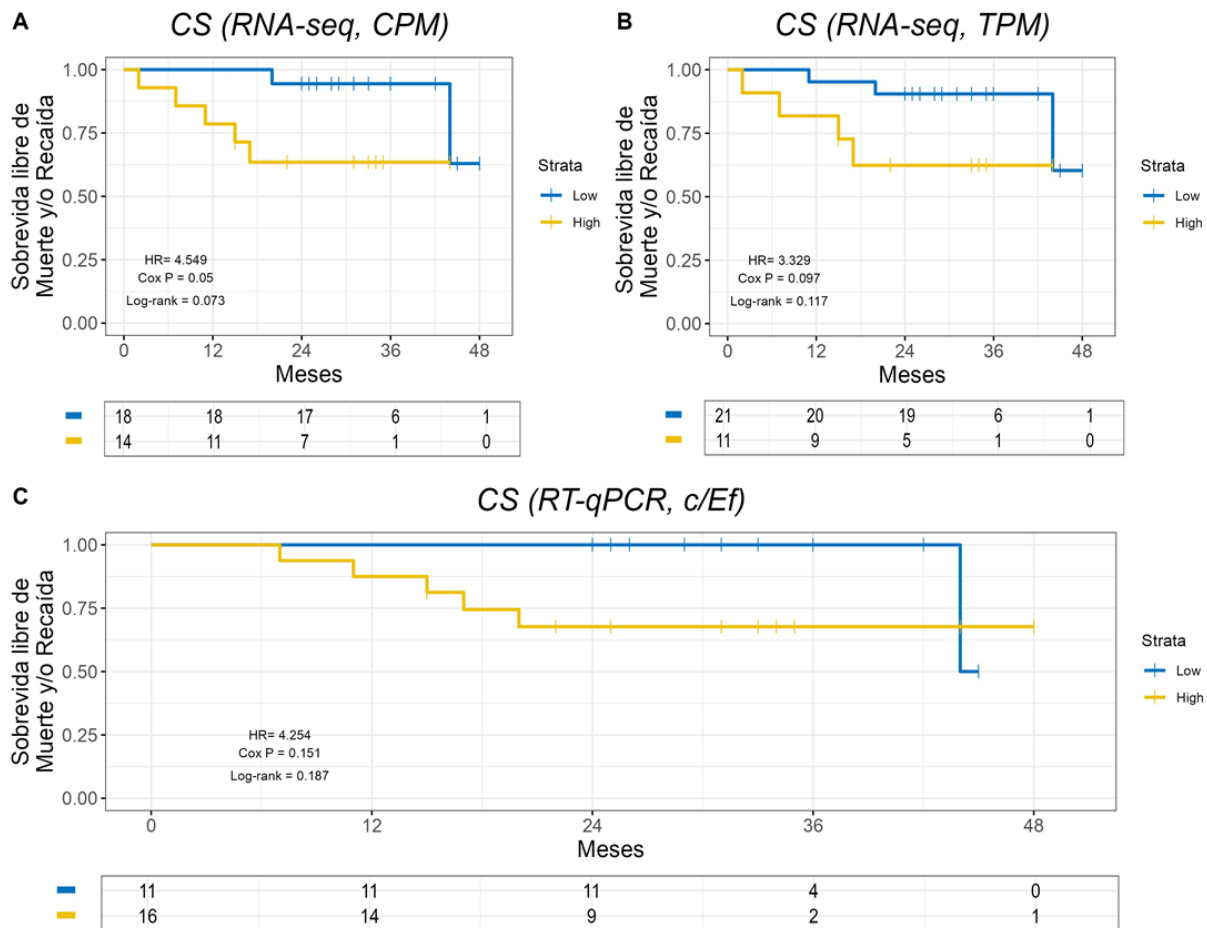
Finalmente, se evaluó si existía una asociación entre la supervivencia global, supervivencia libre de recaída y supervivencia libre de eventos (definido como recaída o muerte, lo primero que ocurra) y el valor del CS mediante curvas de Kaplan-Meier y modelos de regresión de Cox, clasificando a la población de pacientes en CS alto vs. bajo mediante la herramienta *Cutoff Finder* (Andersen et al., 2004). Se observó que los pacientes con CS alto presentaron un mayor riesgo de presentar el evento que aquellos con CS bajo. A pesar de que se detectaron diferencias marginalmente

significativas sólo para el CS estimado por CPM (HR=4.549; p=0,05), en las curvas puede observarse un patrón de menor tiempo al evento, ocurriendo la mayoría antes de los 18 meses en el grupo de CS alto (**figura 25**) (**Tabla 11**).

**Tabla 11: Resultados de curvas de Kaplan Meier dividiendo a los pacientes según su CS (punto de corte estimado por *Cutoff Finder*) utilizando distintos eventos para medir la sobrevida.**

Forma de cuantificación	Sobrevida global			Sobrevida libre de recaída			Sobrevida libre de evento		
	Log-rank	HR	CoxP	Log-rank	HR	CoxP	Log-rank	HR	CoxP
<b>CS (RNA-seq, CPM)</b>	0,098	6,39	<b>0,058</b>	0,279	2,712	0,261	0,073	4,549	<b>0,05</b>
<b>CS (RNA-seq, TPM)</b>	ND	ND	ND	0,489	1,884	0,482	0,117	3,329	0,097
<b>CS (RT-qPCR)</b>	ND	ND	ND	1,772	0,532	0,526	4,254	0,187	0,151

**Tabla 11:** Valores de Hazard Ratio (HR), CoxP y Log-rank de las curvas de sobrevida global, sobrevida libre de recaída y sobrevida libre de evento (definido como recaída o muerte, lo primero que ocurra) para el CS medido por CPM, TPM o RT-qPCR. ND (no determinado): al momento de establecer el punto de corte para separar a las muestras en 2 grupos alguno de los dos no poseía evento, con lo cual el valor resultante carecía de sentido.



**Figura 25: Curvas de Kaplan Meier para la sobrevida libre de evento (muerte o recaída, lo primero que ocurra).** Los pacientes fueron agrupados según su CS estimado por RNaseq en unidades de CPM (A), TPM (B) o por RT-qPCR (C). El valor de corte Low/High fue estimado mediante la herramienta *Cutoff Finder* en los tres casos. Los números en la parte inferior indican el número de pacientes en riesgo en cada punto de tiempo, “Low”: CS bajo (curvas azules); “High”: CS alto (curvas amarillas).

Como conclusión de esta sección, logramos calcular un CS que se correlaciona con la abundancia de células citotóxicas por medio de un CS basado en la expresión de 5 genes que se expresan específicamente en éste tipo de células. Los valores de CS estimados a partir de los datos de RNA-seq presentaron una alta correlación con las proporciones estimadas por MIXTURE de células NK, linfocitos T CD8<sup>+</sup> y T CD4<sup>+</sup> (**figura 19**). Desarrollamos un método de cuantificación del CS basado en RT-qPCR, el cual presentó una alta correlación y concordancia con el CS calculado a partir del RNA-seq, aunque consideramos que requiere validación en un conjunto independiente de muestras (**figura 24**).

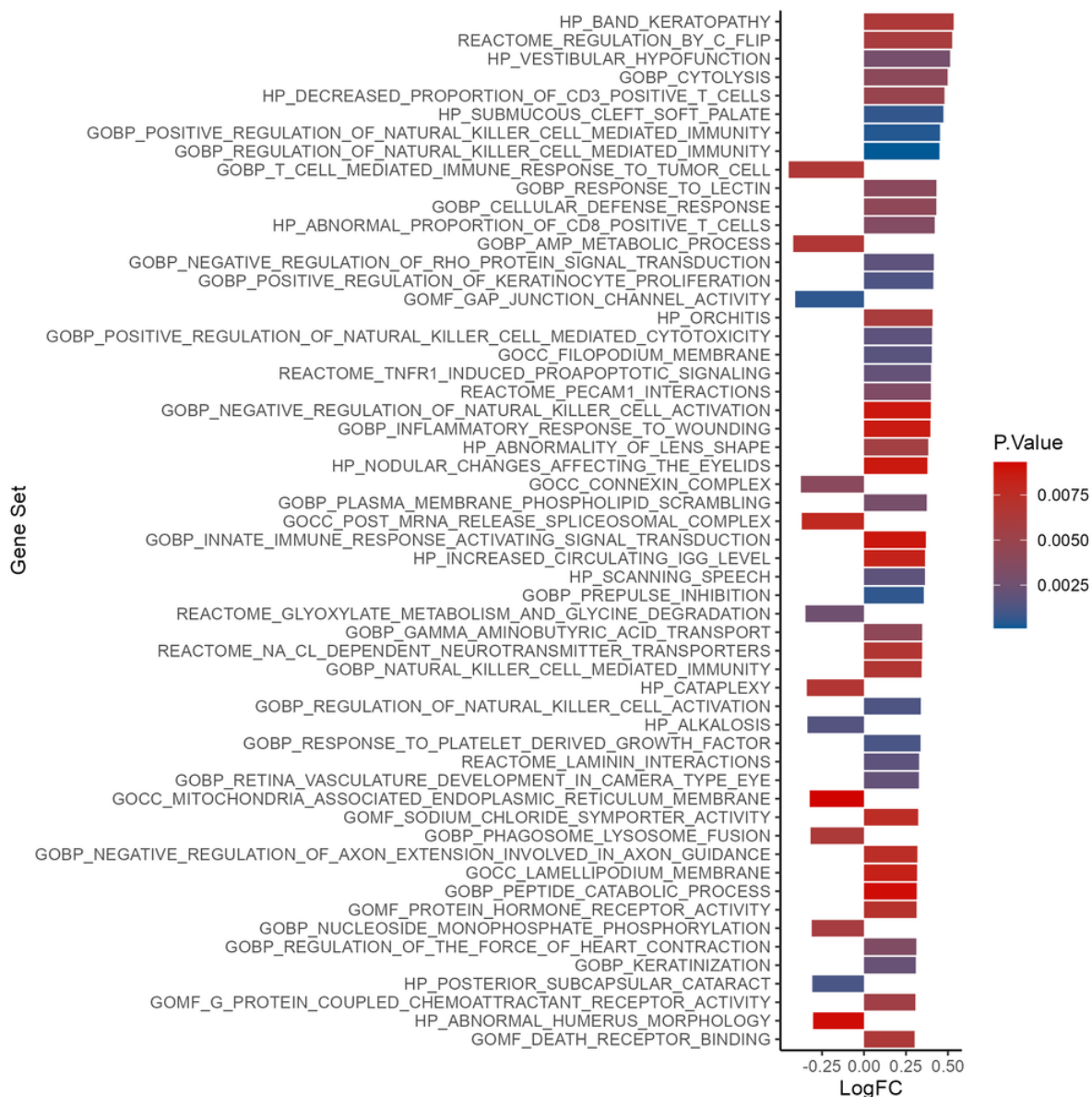
#### IV.4) Caracterización del TME a través de vías biológicas enriquecidas a nivel transcripcional

El análisis realizado por MIXTURE permitió estimar la abundancia de poblaciones inmunitarias. Sin embargo, es de interés explorar qué posibles consecuencias funcionales pueden estar asociadas a su presencia en el TME. Dadas las diferencias observadas en sobrevida libre de muerte y/o recaída para los pacientes con distinto CS se realizó un análisis

de variación de sets de genes (GSVA) y se compararon los puntajes de enriquecimiento entre ambos grupos. Este análisis es un método no paramétrico y no supervisado que permite estimar un índice de enriquecimiento de un set de genes relacionados en un determinado proceso celular) para cada muestra individual dentro de un grupo de muestras; a diferencia de un análisis tradicional por *Gene Set Enrichment Analysis* (GSEA), en el que el índice de enriquecimiento es calculado a partir de una lista de genes diferencialmente expresados entre dos grupos de muestras. Los índices de enriquecimiento obtenidos luego pueden ser utilizados para análisis estadísticos del tipo de modelos lineales generales (Castelo, 2023).

A diferencia de un análisis de expresión diferencial de genes, este método permite evaluar la desregulación de un mismo proceso biológico que puede estar mediado por diferentes genes entre las muestras individuales, lo cual puede ocurrir en muestras con alta variabilidad biológica. A su vez, los índices de enriquecimiento son calculados una única vez para todas las muestras, y luego pueden utilizarse para distintas comparaciones entre grupos, utilizando diversos modelos estadísticos (por ejemplo, modificando las covariables incluidas, o repitiendo el análisis luego de actualizarse el seguimiento clínico). Seleccionamos 7 sets de genes (detallados en Materiales y métodos), y se compararon los valores de índice de enriquecimiento para cada uno de ellos entre pacientes con CS alto (Z-score >1) vs. CS bajo (Z-score <1). Los sets de genes que resultaron significativamente enriquecidos (p-valor sin ajustar <0,01) se muestran en la **figura 26**.

Mediante este análisis se pudo detectar que los pacientes con CS alto poseían un perfil transcriptómico en su TME asociado a: citotoxicidad y citólisis mediada por células NK, inferior proporción de células T CD3<sup>+</sup>, señalización de apoptosis, activación de la respuesta inmune innata, movilización de linfocitos T (PECAM1\_interactions) y regulación de la activación de células NK. Se detectaron también otros sets de genes que, *a priori*, no estarían directamente relacionados con la patología estudiada (como, por ejemplo, proliferación de queratinocitos), pero en los que se podría profundizar en futuros trabajos.



**Figura 26: Genesets con enriquecimiento diferencial entre pacientes con CS alto vs. bajo.** Log<sub>2</sub> Fold Change de los sets de genes más enriquecidos con un p-valor < 0.01. En gradiente de colores se indica el p-valor de cada set de genes. El modelo con el cual se realizó el presente análisis de GSVA contempló como covariables el sexo de los pacientes, el grupo de riesgo al cual pertenecían y el lote de secuenciación de las muestras. Se utilizaron los p-valores sin ajustar ya que los p-valores ajustados no dieron en ningún caso significativos. Este análisis de GSVA se realizó utilizando los datos transcriptómicos en TPM.

Se realizó el mismo análisis de GSVA, pero utilizando los datos transcriptómicos en CPM en lugar de los TPM. Los resultados obtenidos fueron muy similares, aunque, a diferencia del análisis anterior, se destacaron 2 sets de genes relacionados a la presencia de PD1 en linfocitos T CD8<sup>+</sup> (*GSE26495 Naive vs PD1low CD8 T cells DN*) en pacientes con CS alto, sugerente de un fenotipo exhausto.

En ambos análisis realizados no se obtuvieron valores de p-valor ajustados significativos, probablemente debido a los pocos pacientes dentro de la categoría de CS alto. Por este motivo, se realizó otro análisis de GSVA utilizando los datos transcriptómicos en TPM y la separación de pacientes obtenida por el *Cutoff Finder* que puede observarse en la **figura 25 A** con el fin

de aumentar el número de pacientes en la categoría de CS alto. Los resultados fueron similares, aunque para el set de genes *GOBP cytolysis* se obtuvo un p-valor ajustado significativo (p-val= 0,0101). Otros sets de genes relacionados a la citotoxicidad o actividad citolítica fueron: *GOBP immunological synapse formation*, *GOCC cytolytic granule*, *GOBP positive regulation of natural killer cell mediated cytotoxicity*, *REACTOME caspase activation via death receptors in the presence of ligand*. Las observaciones sugieren que podría tratarse de un subgrupo de pacientes con características distintivas de un TME con alta inmunogenicidad, en el cual sería de interés caracterizar la funcionalidad o marcadores de actividad en los linfocitos T CD8<sup>+</sup>, CD4<sup>+</sup> y células NK. En este sentido, como primera aproximación, resultados preliminares del laboratorio sugieren niveles aumentados del marcador de agotamiento TIM-3, del marcador de migración y adherencia celular CX3CR1, y de las citoquinas CCL4 y XCL2 en pacientes con CS alto.

El hecho de que se hayan encontrado sets de genes relacionados a citotoxicidad y citólisis en pacientes con CS alto, en particular mediadas por células NK y señalización de muerte celular, refuerza la aplicabilidad del CS en cuanto a su capacidad de estimar la presencia y actividad de células citotóxicas. Al mismo tiempo, un set de genes relacionado a proporciones disminuidas de células T CD3<sup>+</sup> se vio asociado al grupo de pacientes con CS alto. Este fue un resultado inesperado, dado que el CS, tal como se comprobó al compararlo con las proporciones estimadas por MIXTURE, está asociado a la abundancia de células citotóxicas que incluye a linfocitos T CD8<sup>+</sup> y CD4<sup>+</sup>.



## V) Discusión

Este trabajo de tesis partió de la hipótesis de que parte de la variabilidad clínica observada en los pacientes con LLA puede explicarse por la contribución del componente inmunitario presente en el TME; por lo que nos propusimos caracterizar dicho componente en muestras de pacientes pediátricos argentinos con esta patología. La identificación de factores pronóstico en el TME podría contribuir a mejorar el esquema de estratificación de pacientes en grupos de riesgo, basado actualmente sólo en características de las células leucémicas y en parámetros clínicos al diagnóstico y durante el tratamiento. Adicionalmente, este estudio podría contribuir a identificar nuevos blancos terapéuticos asociados a la interacción entre las células leucémicas y el TME, especialmente en el contexto del desarrollo de inmunoterapias y terapias dirigidas en LLA. En esta tesis, se logró describir el TME de 32 pacientes pediátricos argentinos con LLA-B utilizando datos de expresión génica, mediante citometría digital y por estimación de un índice que refleja la abundancia de células citotóxicas. Los resultados pueden evaluarse desde tres perspectivas: **1) Factibilidad técnica, 2) Descripción del estado del componente inmunitario del TME al momento del diagnóstico en nuestra cohorte de pacientes y 3) Asociación con la respuesta clínica.**

**Respecto a la factibilidad técnica,** la herramienta MIXTURE resultó ser una técnica rápida para el cálculo de las proporciones celulares en base al perfil transcriptómico de una mezcla compleja de células. La abundancia de linfocitos B estimada por ambas firmas correlacionó positivamente con el porcentaje de blastos determinados al diagnóstico mediante una metodología independiente (citomorfología), a la vez que las proporciones de la mayoría de las poblaciones también mostraron una correlación positiva y alta concordancia con las proporciones estimadas por CIBERSORT, una herramienta de deconvolución ampliamente utilizada. Las poblaciones que mostraron discrepancias entre ambas metodologías de citometría digital fueron las células B (estimadas por ambas firmas) y las células T CD4<sup>+</sup> (estimadas por TIL10). Dado que encontramos diferencias significativas en la abundancia de los linfocitos T CD4<sup>+</sup> entre pacientes que presentaron o no muerte y/o recaída, será necesario estimar la abundancia de esta población utilizando otros métodos de deconvolución y/o a partir de otras firmas transcripcionales de referencia, siendo que éstas pueden introducir sesgos en los resultados (Vallania et al., 2018). Por otro lado, las proporciones de células NK, T CD8<sup>+</sup> y T CD4<sup>+</sup> estimadas por MIXTURE presentaron una alta correlación con el CS, el cual ha sido previamente validado en neoplasias hematológicas (Dufva et al., 2020). Al mismo tiempo, el hecho de que las proporciones de las poblaciones celulares para las réplicas disponibles (POS002, POS003 y POS005) hayan presentado una alta concordancia indica una alta reproducibilidad de la técnica. Dado que MIXTURE fue evaluada en datos de TCGA (Tomczak et al., 2015) de cáncer de mama, melanoma, cáncer pulmón y cáncer colorrectal, todos tumores sólidos en donde las células tumorales no son células inmunitarias, en un futuro sería de utilidad el diseño de una firma transcripcional que permita diferenciar entre blastos y células terminalmente diferenciadas.

En conjunto, estos resultados sugieren que MIXTURE es aplicable, al menos, para la estimación de poblaciones inmunitarias del linaje T/NK en LLA de tipo B, aunque hemos

detectado que pueden existir diferencias según la firma utilizada, sugiriendo el análisis con múltiples firmas independientes. A su vez, es importante mencionar que la baja proporción de la mayoría de las poblaciones (nula en algunos casos) debida a la predominancia de células leucémicas en la muestra, fue una limitación de nuestro estudio.

En cuanto al cálculo del CS, se observó una correlación positiva en el cálculo realizado a partir de los datos obtenidos por RNA-seq y los calculados a partir de RT-qPCR. Si bien en el análisis de concordancia entre ambas metodologías por medio de un gráfico de *Bland Altman* se observó que no había una distribución al azar, se debe tener en cuenta que para el cálculo del CS por qPCR no se utilizó el gen *GZMM*. Se tomó esta decisión debido a que los *primers* diseñados no lograron amplificar un producto del tamaño esperado en las muestras y a que existen trabajos que calculan un *cytolytic score* con menos genes, por lo general utilizan únicamente *GZMA* y *PRF1* (Fernández et al., 2021) (Haddad et al., 2020). Esto podría explicar, al menos en parte, que la concordancia no haya sido perfecta. A su vez, el límite de detección de la RT-qPCR fue menor en nuestro ensayo que el del RNA-seq, limitando la comparabilidad de las muestras con menor CS o peor calidad del RNA. La detección mediante sondas de tipo Taqman, la inclusión de un mayor número de genes de referencia y la obtención de curvas de eficiencia con más puntos podrían contribuir a que la concordancia entre ambas metodologías aumente.

**Respecto al estado del componente inmunitario del TME al momento del diagnóstico**, esta patología se caracteriza por presentar una gran heterogeneidad clínica y molecular, y nuestra hipótesis al comienzo de este proyecto fue que el componente inmunitario del TME también presentaba una alta heterogeneidad. Es de interés poder evaluar en el futuro la correlación de este componente con los subtipos moleculares. Al día de hoy se han descrito al menos 18 subtipos moleculares en LLA-B (Gu et al., 2019), los cuales están asociados a una evolución clínica diferencial. Por otro lado, a nuestro saber, este es el primer reporte en pacientes pediátricos con LLA-B que aplique esta metodología. Mediante citometría de flujo, un estudio en 100 pacientes pediátricos de LLA-B reportó una mayor proporción de células T efectoras en estadios tardíos (CD62L<sup>-</sup>/CD45RO<sup>+</sup>) en la médula ósea en comparación con individuos sanos (Blaeschke et al., 2020). Otro estudio describió la emergencia de monocitos no clásicos al momento del diagnóstico y durante la recaída, a la par que la abundancia de monocitos está relacionada a una menor sobrevida libre de recaída y sobrevida global, tanto en el microambiente tumoral de LLA-B adulta como pediátrica (Witkowski et al., 2020).

La comparación del perfil obtenido en los pacientes con LLA-B con muestras de médula ósea de donadores sanos arrojó resultados interesantes para seguir estudiando. En el contexto del resto de los resultados obtenidos en este trabajo, nos resultó de particular interés observar que la proporción de células NK fue similar entre donadores sanos y LLA-B. Siendo que en pacientes se espera una proporción menor de las poblaciones celulares dada la expansión de células del linaje B, podemos plantear la hipótesis de que el número absoluto de células NK se encuentra aumentado en la médula ósea de pacientes con LLA-B al momento del diagnóstico.

Otra población celular interesante que se destacó a lo largo de los distintos análisis realizados en esta tesis fueron los linfocitos T CD4<sup>+</sup>. Al realizar agrupamientos no supervisados se

observó que el *cluster* 3 (tanto en los dendrogramas realizados con las proporciones estimadas por TIL10 como LM22) se diferenciaba de los otros dos clusters por una menor proporción de esta población celular. También se encontró que los genes del CS se expresaban en al menos una subpoblación de esta población celular (*Human Protein Atlas*) y que el CS calculado con los datos de RNA-seq correlacionaba positivamente con la proporción de esta población (MIXTURE, TIL10). Estudios recientes han demostrado que la actividad citotóxica no es exclusiva de células NK y linfocitos T CD8<sup>+</sup>, sino que ciertas subpoblaciones de linfocitos T CD4<sup>+</sup> también pueden ejercer esta actividad. Estudios *in vivo* en un modelo de melanoma encontraron que linfocitos T CD4<sup>+</sup> adquirieron actividad citotóxica pudiendo eliminar células tumorales que expresan moléculas del MHC de clase II (Quezada et al., 2010). Otro estudio reportó que linfocitos T CD4<sup>+</sup> *naive* podían diferenciarse a linfocitos T *helper* de tipo 1 con actividad citotóxica (Xie et al., 2010). Śledzińska et al. demostraron que esta capacidad citotóxica se adquiere con la expresión de los factores de transcripción Blimp-1 y T-bet (Śledzińska et al., 2020). En este aspecto, estas células poseen un potencial sub-explorado como blancos terapéuticos o en inmunoterapias.

**Respecto a la asociación con la respuesta clínica**, existen dos limitantes a las conclusiones que se pueden sacar de este estudio. Por un lado, la mediana de tiempo de seguimiento de los pacientes fue de 32 meses (rango intercuartil: 25-35) lo cual, combinado al número total de pacientes reclutados, resultó en un bajo número de pacientes en cada categoría clínica, limitando la potencia del estudio. Proponemos dos estrategias para mejorar este aspecto: por un lado, el CS medido RT-qPCR será calculado en un grupo de 32 muestras de LLA-B extras para las cuales no contamos con datos de RNA-seq pero tenemos los RNA. Por otro lado, se aplicará la misma metodología de citometría digital y CS a un conjunto de datos transcriptómicos provenientes de 288 pacientes con LLA-B disponibles en repositorios de acceso público (cohorte TARGET), para evaluar si existe concordancia en los patrones observados. Es importante tener en cuenta también que los pacientes analizados en este trabajo de tesis fueron sometidos a un tratamiento de poliquimioterapia. Por lo tanto, no podemos descartar que el componente inmunitario del TME pueda ser relevante en la respuesta clínica a otras terapias, como pueden ser la inmunoterapia y la terapia dirigida (ej. inhibidores de tirosín-quinasa, algunos de los cuales han sido descriptos con acción en el sistema inmunitario (Mustjoki et al., 2013) (Climent & Plana, 2019) ). En este sentido, esperamos que la profundización de este estudio permita proponer marcadores para su futuro estudio en ensayos clínicos en los cuales se utilicen estrategias terapéuticas alternativas, como así también proponer nuevas hipótesis de interacción entre el sistema inmunitario y las células leucémicas que puedan ser testeadas en ensayos en líneas celulares y modelos animales.

Aún con estas limitaciones, se encontró una asociación entre la abundancia de linfocitos T CD4<sup>+</sup> y monocitos con muerte y/o recaída; y entre la abundancia de linfocitos T CD8<sup>+</sup> y células NK con toxicidad aguda relacionada al tratamiento (**tabla 6**). Se requerirán más estudios para validar esta observación y entender el origen de dichas asociaciones.

Se han descripto casos en los que un TME pro-inflamatorio puede promover el crecimiento tumoral y estar asociado a peor pronóstico. Tal es el caso de los monocitos no-clásicos, los

cuales se caracterizan por poseer una actividad pro-inflamatoria y que en un estudio reciente se halló una asociación entre estas células y una menor sobrevida libre de recaída en LLA-B (Witkowski et al., 2020). Si bien ninguna de las firmas utilizadas en esta tesis permite distinguir entre subtipos de monocitos, es de interés profundizar en la asociación hallada entre monocitos y la sobrevida. Hallazgos recientes sugieren que un microambiente inflamatorio en la médula ósea puede influir en el proceso de transformación maligna de precursores pre-leucémicos (Isidro-Hernández et al., 2020). En este trabajo se describió que, en comparación con líneas celulares proB *wild type*, células proB *Pax5<sup>+/-</sup>* (un gen que se encuentra mutado en algunos casos de LLA-B) producen mayores niveles de interleuquina-6 (una citocina inflamatoria), la cual tiene función leucemogénica en este modelo. En otro estudio se encontró una disminución en la expresión de receptores activadores en las células NK de pacientes pediátricos de LLA en comparación a controles sanos de la misma (Valenzuela-Vázquez et al., 2022).

En cuanto a la asociación entre el CS y las variables clínicas, está reportado que la actividad citolítica está relacionada con una mejor sobrevida en estudios de pan-cancer (Rooney et al., 2015) y en varios tumores sólidos (Roufas et al., 2018). Se ha reportado una asociación entre altos valores de CS, medido a través de GZMA y PRF1, y baja heterogeneidad intratumoral en melanoma (Fernández et al., 2021). En este tipo de cáncer, una baja heterogeneidad intratumoral está asociada a una mejor sobrevida. Por el contrario, en glioblastoma se reportó que un alto CS (basado en la expresión de los mismos genes) estaba asociado a un peor pronóstico (Haddad et al., 2020); sugiriendo que la abundancia de células citotóxicas tiene consecuencias muy diferentes en distintos tipos de cánceres. En esta tesis no se detectaron diferencias estadísticamente significativas en la sobrevida global o sobrevida libre de muerte/recaída con respecto a los valores del CS (medido tanto por RT-qPCR como por RNA-seq). Sin embargo, el análisis por curvas de sobrevida evidenció una tendencia similar a la reportada en glioblastoma, con valores mayores de CS asociados a un mayor riesgo de muerte/recaída, que es de interés profundizar en futuros estudios.

El estado de agotamiento o fenotipo exhausto es un término amplio que ha sido utilizado para describir la respuesta de células T a la estimulación antigénica crónica, como la que puede darse en infecciones virales crónicas o en respuesta a tumores. La caracterización de este estado funcional en el TME podría contribuir a entender las interacciones entre las células tumorales y el componente inmunitario. Este fenotipo exhausto se caracteriza por la expresión de receptores inhibitorios como PD-1, CTLA4, TIM-3 y LAG-3, lo que conduce a una actividad efectora disminuida de estas células, pudiendo contribuir al escape tumoral. En un trabajo se reportó una expresión elevada de estos marcadores en pacientes de LLA-B que experimentaron recaída luego de un trasplante de células madre hematopoyéticas alogénicas (Liu et al., 2018). En otro trabajo reciente se describieron, mediante *single-cell* RNA-seq, poblaciones de células T exhaustas en sangre periférica de pacientes con LLA-B (Wang et al., 2021). Los autores explican que estas poblaciones eran muy heterogéneas en cuanto a su expresión de *GNLY*, *PRF1* y *GZMH*, entre otras características, pudiendo diferenciar distintos subgrupos. En el trabajo especulan que el subgrupo que mantiene cierta actividad citotóxica puede deberse a un grupo de células T *naive* en el proceso de convertirse en células T exhaustas. Esto nos lleva a plantear una nueva hipótesis en la que la asociación entre una mayor abundancia de células

citotóxicas y menor sobrevida está relacionada con un fenotipo exhausto de estas células. Sin embargo, dada la complejidad del tratamiento y la heterogeneidad genética y clínica de esta patología, estas observaciones requieren una validación en un grupo mayor de pacientes para poder controlar otras variables con valor pronóstico en el análisis de sobrevida.

Se reportó que la presencia del marcador de agotamiento TIM-3 en células T CD4<sup>+</sup> estaba asociada a un mayor riesgo de recaída y/o muerte en pacientes pediátricos de LLA-B (Blaeschke et al., 2020). En pacientes adultos con LLA, la proporción de linfocitos T CD4<sup>+</sup>PD1<sup>+</sup>TIM3<sup>+</sup>, medidos por inmunohistoquímica multiparamétrica en médula ósea, fue un factor pronóstico independiente en un modelo multivariable, diferenciando un grupo con peor sobrevida (Hohtari et al., 2019). En el mismo estudio se reportó una asociación entre la disminución de la proporción de células T activadas y macrófagos tipo M1, así como en una actividad citolítica reducida, y elevados niveles de macrófagos M2 y células supresoras derivadas de mieloides (MDSCs) en la médula ósea de LLA, lo cual sugiere que un microambiente tumoral inmunosupresor está asociado al desarrollo de la leucemia.

Finalmente, la exploración de las vías biológicas enriquecidas en pacientes con CS alto vs. bajo permitió confirmar por una metodología alternativa la presencia y actividad de células del linaje linfóide T/NK citotóxicas en el primer grupo. Estos resultados sugieren que el CS permite distinguir en los pacientes la inmunidad mediada por este tipo de células. Una de las limitaciones al realizar este análisis es el hecho de que la definición del CS “alto” o “bajo” es relativo a esta cohorte de pacientes, y que el grupo de CS alto estaba compuesto por sólo 4 pacientes. En este sentido, sería importante poder asociar este parámetro al número de células T/NK en médula ósea por una metodología independiente (ej. citometría de flujo). Sería de interés también poder estudiar la composición del TME a distintos tiempos de iniciado el tratamiento para evaluar los cambios inducidos por la quimioterapia. En concordancia con lo descrito en bibliografía (Dufva et al., 2020), pudimos identificar un subgrupo minoritario de pacientes con una mayor abundancia de células citotóxicas en la médula ósea en el momento de diagnóstico, lo que es sugerente de un mayor reclutamiento y activación de la respuesta inmunitaria en estos pacientes. En este sentido, en Argentina se está empezando a utilizar rituximab de forma randomizada como parte del protocolo clínico multicéntrico ALLIC-BFM 2022 y blinatumomab como parte de las guías de tratamiento para pacientes en recaída (GATLA-LLA-REC-2022). Por lo que los resultados aquí presentados podrían ser relevantes en futuros estudios que busquen identificar pacientes con respuesta diferencial a estas inmunoterapias.

En resumen, en esta tesis se estimaron las proporciones de las poblaciones celulares que conforman el componente inmunitario del TME de pacientes argentinos pediátricos con LLA utilizando la herramienta de deconvolución MIXTURE. Se detectaron niveles significativamente menores de monocitos y mayores de células T CD4<sup>+</sup> en pacientes que recayeron o murieron, y niveles significativamente mayores de células T CD8<sup>+</sup> y NK en pacientes que desarrollaron toxicidad aguda relacionada al tratamiento. Se evaluó el desempeño de un índice capaz de identificar pacientes con mayor abundancia y actividad de células citotóxicas (NK, T CD4<sup>+</sup> y T CD8<sup>+</sup>) en base a la expresión de 5 genes, y se desarrolló

un ensayo de RT-qPCR para su cuantificación en muestra de pacientes. Se observó un patrón (aunque las diferencias no fueron estadísticamente significativas) en el que pacientes con un CS alto presentaron una menor sobrevida libre de muerte y/o recaída, y un perfil transcripcional concordante con la activación, y tal vez agotamiento, de las células citotóxicas.

## VI) Bibliografía

- Abbas, A.K., Lichtman, A.H. and Pillai, S. (2015). *Cellular and Molecular Immunology*. 8th Edition.
- Abbas, A. R., Wolslegel, K., Seshasayee, D., Modrusan, Z., & Clark, H. F. (2009). Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS ONE*, 4(7). <https://doi.org/10.1371/journal.pone.0006098>
- Abbate, M. (2023). *Estudio de mecanismos moleculares involucrados en el desarrollo y progresión de la leucemia linfoblástica aguda en pediatría*.
- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., ... Westerfield, M. (2023). The Gene Ontology Knowledgebase in 2023. *Genetics*, 224(1), 1–14. <https://doi.org/10.1093/genetics/iyad031>
- Altman, J. M. B. D. G. (1986). MEASUREMENT, STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL. *The Lancet*. [https://doi.org/https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/https://doi.org/10.1016/S0140-6736(86)90837-8)
- American Cancer Society, I. (2023). *2023-Cancer-Facts-and-Figures*.
- Andersen, C. L., Jensen, J. L., & Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research*, 64(15), 5245–5250. <https://doi.org/10.1158/0008-5472.CAN-04-0496>
- Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology*, 30(16), R921–R925. <https://doi.org/10.1016/j.cub.2020.06.081>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Blaeschke, F., Willier, S., Stenger, D., Lepenies, M., Horstmann, M. A., Escherich, G., Zimmermann, M., Rojas Ringeling, F., Canzar, S., Kaeuferle, T., Rohlf, M., Binder, V., Klein, C., & Feuchtinger, T. (2020). Leukemia-induced dysfunctional TIM-3+CD4+ bone marrow T cells increase risk of relapse in pediatric B-precursor ALL patients. *Leukemia*, 34(10), 2607–2620. <https://doi.org/10.1038/s41375-020-0793-1>
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5), 525–527. <https://doi.org/10.1038/nbt.3519>

- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, 9(2), 378–400. <https://doi.org/10.32614/rj-2017-066>
- Budczies, J., Klauschen, F., Sinn, B. V., Györffy, B., Schmitt, W. D., Darb-Esfahani, S., & Denkert, C. (2012). Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization. *PLoS ONE*, 7(12), 1–7. <https://doi.org/10.1371/journal.pone.0051862>
- Budczies, J., Klauschen, F., Sinn, B. V., Györffy, B., Schmitt, W. D., Darb-Esfahani, S., & Denkert, C. (2012). Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PloS One*, 7(12), e51862. <https://doi.org/10.1371/journal.pone.0051862>
- Castelo, M. R. (2023). *Package ‘GSVA.’* <https://github.com/rcastelo/GSVA>
- Climent, N., & Plana, M. (2019). Immunomodulatory activity of tyrosine kinase inhibitors to elicit cytotoxicity against cancer and viral infection. *Frontiers in Pharmacology*, 10(October), 1–9. <https://doi.org/10.3389/fphar.2019.01232>
- Cullen, S. P., Brunet, M., & Martin, S. J. (2010). Granzymes in cancer and immunity. In *Cell Death and Differentiation* (Vol. 17, Issue 4, pp. 616–623). <https://doi.org/10.1038/cdd.2009.206>
- Denton, A. E., Roberts, E. W., & Fearon, D. T. (2018). Stromal Cells in the Tumor Microenvironment. In B. M. J. Owens & M. A. Lakins (Eds.), *Stromal Immunology* (pp. 99–114). Springer International Publishing. [https://doi.org/10.1007/978-3-319-78127-3\\_6](https://doi.org/10.1007/978-3-319-78127-3_6)
- Dufva, O., Pölönen, P., Brück, O., Keränen, M. A. I., Klievink, J., Mehtonen, J., Huuhtanen, J., Kumar, A., Malani, D., Siitonen, S., Kankainen, M., Ghimire, B., Lahtela, J., Mattila, P., Vähä-Koskela, M., Wennerberg, K., Granberg, K., Leivonen, S. K., Meriranta, L., ... Mustjoki, S. (2020). Immunogenomic Landscape of Hematological Malignancies. *Cancer Cell*, 38(3), 380–399.e13. <https://doi.org/10.1016/j.ccell.2020.06.002>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Feola, S., Chiaro, J., Martins, B., & Cerullo, V. (2020). Uncovering the tumor antigen landscape: What to know about the discovery process. *Cancers*, 12(6), 1–28. <https://doi.org/10.3390/cancers12061660>
- Fernández, E. A., Mahmoud, Y. D., Veigas, F., Rocha, D., Miranda, M., Merlo, J., Balzarini, M., Lujan, H. D., Rabinovich, G. A., & Girotti, M. R. (2021). Unveiling the immune infiltrate modulation in cancer and response to immunotherapy by MIXTURE-an enhanced deconvolution method. *Briefings in Bioinformatics*, 22(4). <https://doi.org/10.1093/bib/bbaa317>



- Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., Krogsdam, A., Loncova, Z., Posch, W., Wilflingseder, D., Sopper, S., Ijsselsteijn, M., Brouwer, T. P., Johnson, D., Xu, Y., Wang, Y., Sanders, M. E., Estrada, M. V., Ericsson-Gonzalez, P., ... Trajanoski, Z. (2019). Correction to: Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data (*Genome Medicine* (2019) 11 (34) DOI: 10.1186/s13073-019-0638-6). *Genome Medicine*, 11(1), 1–20. <https://doi.org/10.1186/s13073-019-0655-5>
- Galili, T., O’Callaghan, A., Sidi, J., & Sievert, C. (2018). Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9), 1600–1602. <https://doi.org/10.1093/bioinformatics/btx657>
- Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., ... D’Eustachio, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1), D687–D692. <https://doi.org/10.1093/nar/gkab1028>
- Gizem Sonugür, F., & Akbulut, H. (2019). The role of tumor microenvironment in genomic instability of malignant tumors. *Frontiers in Genetics*, 10(OCT), 1–7. <https://doi.org/10.3389/fgene.2019.01063>
- Globocan. (2020). Cancer Today. In *International Agency for research* (Vol. 418, pp. 1–2) Fecha de acceso: abril 2023. [https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode\\_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=1&statistic=5&prevalence=0&population\\_group=0&ages\\_group%5B%5D=0&ages\\_group%5B%5D=17&group\\_cancer=1&include\\_nmsc=0&include\\_nmsc\\_other=1](https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=1&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=0&include_nmsc_other=1)
- Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A. J., Mesirov, J. P., & Haining, W. N. (2016). Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity*, 44(1), 194–206. <https://doi.org/10.1016/j.immuni.2015.12.006>
- Gonzalez, H., Hagerling, C., & Werb, Z. (2018). Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes and Development*, 32(19–20), 1267–1284. <https://doi.org/10.1101/GAD.314617.118>
- Gu, Z., Churchman, M. L., Roberts, K. G., Moore, I., Zhou, X., Nakitandwe, J., Hagiwara, K., Pelletier, S., Gingras, S., Berns, H., Payne-Turner, D., Hill, A., Iacobucci, I., Shi, L., Pounds, S., Cheng, C., Pei, D., Qu, C., Newman, S., ... Mullighan, C. G. (2019). PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nature Genetics*, 51(2), 296–307. <https://doi.org/10.1038/s41588-018-0315-5>
- Haddad, A. F., Chen, J. S., Oh, T., Pereira, M. P., Joshi, R. S., & Aghi, M. K. (2020). Higher cytolytic score correlates with an immunosuppressive tumor microenvironment and reduced survival in glioblastoma. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-73793-8>
- Hohtari, H., Brück, O., Blom, S., Turkki, R., Sinisalo, M., Kovanen, P. E., Kallioniemi, O.,

- Pellinen, T., Porkka, K., & Mustjoki, S. (2019). Immune cell constitution in bone marrow microenvironment predicts outcome in adult ALL. *Leukemia*, *33*(7), 1570–1582. <https://doi.org/10.1038/s41375-018-0360-1>
- Isidro-Hernández, M., Mayado, A., Casado-García, A., Martínez-Cano, J., Palmi, C., Fazio, G., Orfao, A., Ribera, J., Ribera, J. M., Zamora, L., Raboso-Gallego, J., Blanco, O., Alonso-López, D., De Las Rivas, J., Jiménez, R., García Criado, F. J., García Cenador, M. B., Ramírez-Orellana, M., Cazzaniga, G., ... Sánchez-García, I. (2020). Inhibition of inflammatory signaling in Pax5 mutant cells mitigates B-cell leukemogenesis. *Scientific Reports*, *10*(1), 1–14. <https://doi.org/10.1038/s41598-020-76206-y>
- Jabbour, E., O'Brien, S., Ravandi, F., & Kantarjian, H. (2015). Monoclonal antibodies in acute lymphoblastic leukemia. *Blood*, *125*(26), 4010–4016. <https://doi.org/10.1182/blood-2014-08-596403>
- Jagannathan-Bogdan, M., & Zon, L. I. (2013). Hematopoiesis. *Development (Cambridge)*, *140*(12), 2463–2467. <https://doi.org/10.1242/dev.083147>
- Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M., Callahan, T. J., Chute, C. G., Est, J. L., Galer, P. D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., ... Robinson, P. N. (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, *49*(D1), D1207–D1217. <https://doi.org/10.1093/nar/gkaa1043>
- Lee, J. Y., & Hong, S. H. (2020). Hematopoietic stem cells and their roles in tissue regeneration. *International Journal of Stem Cells*, *13*(1), 1–12. <https://doi.org/10.15283/ijsc19127>
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. <https://doi.org/10.1093/bioinformatics/bts034>
- Li, X., Yang, Y., Huang, Q., Deng, Y., Guo, F., Wang, G., & Liu, M. (2021). Crosstalk Between the Tumor Microenvironment and Cancer Cells: A Promising Predictive Biomarker for Immune Checkpoint Inhibitors. *Frontiers in Cell and Developmental Biology*, *9*(October). <https://doi.org/10.3389/fcell.2021.738373>
- Liberzon, A., Birger, C., Ghandi, M., Jill, P., Tamayo, P., Jolla, L., & Jolla, L. (2016). *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. *1*(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- Lilljebjörn, H., Henningson, R., Hyrenius-Wittsten, A., Olsson, L., Orsmark-Pietras, C., Von Palffy, S., Askmyr, M., Rissler, M., Schrappe, M., Cario, G., Castor, A., Pronk, C. J. H., Behrendtz, M., Mitelman, F., Johansson, B., Paulsson, K., Andersson, A. K., Fontes, M., & Fioretos, T. (2016). Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nature Communications*, *7*. <https://doi.org/10.1038/ncomms11790>
- Liu, L., Chang, Y. J., Xu, L. P., Zhang, X. H., Wang, Y., Liu, K. Y., & Huang, X. J. (2018). T cell exhaustion characterized by compromised MHC class I and II restricted cytotoxic

- activity associates with acute B lymphoblastic leukemia relapse after allogeneic hematopoietic stem cell transplantation. *Clinical Immunology*, 190, 32–40. <https://doi.org/10.1016/j.clim.2018.02.009>
- Lowe, S. W., Cepero, E., & Evan, G. (2004). Intrinsic tumour suppression. *Nature*, 432(7015), 307–315. <https://doi.org/10.1038/nature03098>
- Malagola, M., Papayannidis, C., & Baccarani, M. (2016). Tyrosine kinase inhibitors in Ph+ acute lymphoblastic leukaemia: facts and perspectives. *Annals of Hematology*, 95(5), 681–693. <https://doi.org/10.1007/s00277-016-2617-y>
- Malard, F., & Mohty, M. (2020). Acute lymphoblastic leukaemia. *The Lancet*, 395(10230), 1146–1162. [https://doi.org/10.1016/S0140-6736\(19\)33018-1](https://doi.org/10.1016/S0140-6736(19)33018-1)
- McNeer, J. L., Rau, R. E., Gupta, S., Maude, S. L., & O'Brien, M. M. (2020). Cutting to the Front of the Line: Immunotherapy for Childhood Acute Lymphoblastic Leukemia. *American Society of Clinical Oncology Educational Book*, 40, e132–e143. [https://doi.org/10.1200/edbk\\_278171](https://doi.org/10.1200/edbk_278171)
- Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y. Y., Carré, C., Burdin, N., Visan, L., Ceccarelli, M., Poidinger, M., Zippelius, A., Pedro de Magalhães, J., & Larbi, A. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports*, 26(6), 1627-1640.e7. <https://doi.org/10.1016/j.celrep.2019.01.041>
- Moreno, F., & Chaplin, M. A. (2021). *Registro Oncopediátrico Hospitalario Argentino. Resultados 2000-2019*. <http://www.msal.gob.ar/images/stories/bes/graficos/0000001365cnt-registro-oncopediatico-argentino-digital.pdf>
- Mustjoki, S., Auvinen, K., Kreutzman, A., Rousselot, P., Hernesniemi, S., Melo, T., Lahesmaa-Korpinen, A.-M., Hautaniemi, S., Bouchet, S., Molimard, M., Smykla, R., Lee, F. Y., Vakkila, J., Jalkanen, S., Salmi, M., & Porkka, K. (2013). Rapid mobilization of cytotoxic lymphocytes induced by dasatinib therapy. *Leukemia*, 27(4), 914–924. <https://doi.org/10.1038/leu.2012.348>
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M., & Alizadeh, A. A. (2015a). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, 12(5), 453–457. <https://doi.org/10.1038/nmeth.3337>
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M., & Alizadeh, A. A. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature Biotechnology*, 37(7), 773–782. <https://doi.org/10.1038/s41587-019-0114-2>
- Peña-Romero, A. C., & Orenes-Piñero, E. (2022). Dual Effect of Immune Cells within Tumour Microenvironment: Pro- and Anti-Tumour Effects and Their Triggers. *Cancers*, 14(7). <https://doi.org/10.3390/cancers14071681>

- Pieters, R., & Carroll, W. L. (2008). Biology and Treatment of Acute Lymphoblastic Leukemia. *Pediatric Clinics of North America*, 55(1), 1–20. <https://doi.org/10.1016/j.pcl.2007.11.002>
- Pui, C. H., Nichols, K. E., & Yang, J. J. (2019). Somatic and germline genomics in paediatric acute lymphoblastic leukaemia. *Nature Reviews Clinical Oncology*, 16(4), 227–240. <https://doi.org/10.1038/s41571-018-0136-6>
- Quezada, S. A., Simpson, T. R., Peggs, K. S., Merghoub, T., Vider, J., Fan, X., Blasberg, R., Yagita, H., Muranski, P., Antony, P. A., Restifo, N. P., & Allison, J. P. (2010). Tumor-reactive CD4+ T cells develop cytotoxic activity and eradicate large established melanoma after transfer into lymphopenic hosts. *Journal of Experimental Medicine*, 207(3), 637–650. <https://doi.org/10.1084/jem.20091918>
- Roberts, K. G., Pei, D., Campana, D., Payne-Turner, D., Li, Y., Cheng, C., Sandlund, J. T., Jeha, S., Easton, J., Becksfort, J., Zhang, J., Coustan-Smith, E., Raimondi, S. C., Leung, W. H., Relling, M. V., Evans, W. E., Downing, J. R., Mullighan, C. G., & Pui, C. H. (2014). Outcomes of children with BCR-ABL1-like acute lymphoblastic leukemia treated with risk-directed therapy based on the levels of minimal residual disease. *Journal of Clinical Oncology*, 32(27), 3012–3020. <https://doi.org/10.1200/JCO.2014.55.4105>
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., & Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*, 160(1–2), 48–61. <https://doi.org/10.1016/j.cell.2014.12.033>
- Roufas, C., Chasiotis, D., Makris, A., Efstathiades, C., Dimopoulos, C., & Zaravinos, A. (2018). The expression and prognostic impact of immune cytolytic activity-related markers in human malignancies: A comprehensive meta-analysis. *Frontiers in Oncology*, 8(FEB), 1–18. <https://doi.org/10.3389/fonc.2018.00027>
- Schilstra, C. E., McCleary, K., Fardell, J. E., Donoghoe, M. W., McCormack, E., Kotecha, R. S., Lourenco, R. D. A., Ramachandran, S., Cockcroft, R., Conyers, R., Cross, S., Dalla-Pozza, L., Downie, P., Revesz, T., Osborn, M., Alvaro, F., Wakefield, C. E., Marshall, G. M., Mateos, M. K., & Trahair, T. N. (2022). Prospective longitudinal evaluation of treatment-related toxicity and health-related quality of life during the first year of treatment for pediatric acute lymphoblastic leukemia. *BMC Cancer*, 22(1), 1–12. <https://doi.org/10.1186/s12885-022-10072-x>
- Śledzińska, A., Vila de Mucha, M., Bergerhoff, K., Hotblack, A., Demane, D. F., Ghorani, E., Akarca, A. U., Marzolini, M. A. V., Solomon, I., Vargas, F. A., Pule, M., Ono, M., Seddon, B., Kassiotis, G., Ariyan, C. E., Korn, T., Marafioti, T., Lord, G. M., Stauss, H., ... Quezada, S. A. (2020). Regulatory T Cells Restrain Interleukin-2- and Blimp-1-Dependent Acquisition of Cytotoxic Function by CD4+ T Cells. *Immunity*, 52(1), 151–166.e6. <https://doi.org/10.1016/j.immuni.2019.12.007>
- Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 1–19. <https://doi.org/10.12688/F1000RESEARCH.7563.2>
- Stanulla, M., Cavé, H., & Moorman, A. V. (2020). IKZF1 deletions in pediatric acute

- lymphoblastic leukemia: Still a poor prognostic marker? *Blood*, 135(4), 252–260. <https://doi.org/10.1182/blood.2019000813>
- Taylor, S. C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., & Fenrich, J. (2019). The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology*, 37(7), 761–774. <https://doi.org/10.1016/j.tibtech.2018.12.002>
- Therneau, T. (2021). A Package for Survival Analysis in R; Version 3.2-11. ., <https://CRAN.R-project.org/package=survival>. <https://cran.r-project.org/package=survival>
- Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). ReviewThe Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia*, 68–77. <https://doi.org/10.5114/wo.2014.47136>
- Tuong, P. N., Kiem Hao, T., & Kim Hoa, N. T. (2020). Relapsed Childhood Acute Lymphoblastic Leukemia: A Single-Institution Experience. *Cureus*, 12(7), 1–7. <https://doi.org/10.7759/cureus.9238>
- Valenzuela-Vázquez, L., Nuñez-Enriquez, J. C., Sánchez-Herrera, J., Medina-Sanson, A., Pérez-Saldivar, M. L., Jiménez-Hernández, E., Martiín-Trejo, J. A., Del Campo-Martínez, M. de L. Á., Flores-Lujano, J., Amador-Sánchez, R., Mora-Ríos, F. G., Peñaloza-González, J. G., Duarte-Rodríguez, D. A., Torres-Nava, J. R., Espinosa-Elizondo, R. M., Cortés-Herrera, B., Flores-Villegas, L. V., Merino-Pasaye, L. E., Almeida-Hernández, C., ... Cruz-Munoz, M. E. (2022). NK cells with decreased expression of multiple activating receptors is a dominant phenotype in pediatric patients with acute lymphoblastic leukemia. *Frontiers in Oncology*, 12(November), 1–15. <https://doi.org/10.3389/fonc.2022.1023510>
- Vallania, F., Tam, A., Lofgren, S., Schaffert, S., Azad, T. D., Bongen, E., Haynes, W., Alsup, M., Alonso, M., Davis, M., Engleman, E., & Khatri, P. (2018). Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-07242-6>
- Vora, A., Goulden, N., Wade, R., Mitchell, C., Hancock, J., Hough, R., Rowntree, C., & Richards, S. (2013). Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): A randomised controlled trial. *The Lancet Oncology*, 14(3), 199–209. [https://doi.org/10.1016/S1470-2045\(12\)70600-9](https://doi.org/10.1016/S1470-2045(12)70600-9)
- Wang, X., Chen, Y., Li, Z., Huang, B., Xu, L., Lai, J., Lu, Y., Zha, X., Liu, B., Lan, Y., & Li, Y. (2021). Single-Cell RNA-Seq of T Cells in B-ALL Patients Reveals an Exhausted Subset with Remarkable Heterogeneity. *Advanced Science*, 8(19), 1–12. <https://doi.org/10.1002/advs.202101447>
- Wherry, E. J., & Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology*, 15(8), 486–499. <https://doi.org/10.1038/nri3862>
- Wickham, H. (2011). Ggplot2: Create Elegant Data Visualisations Using the Grammar of

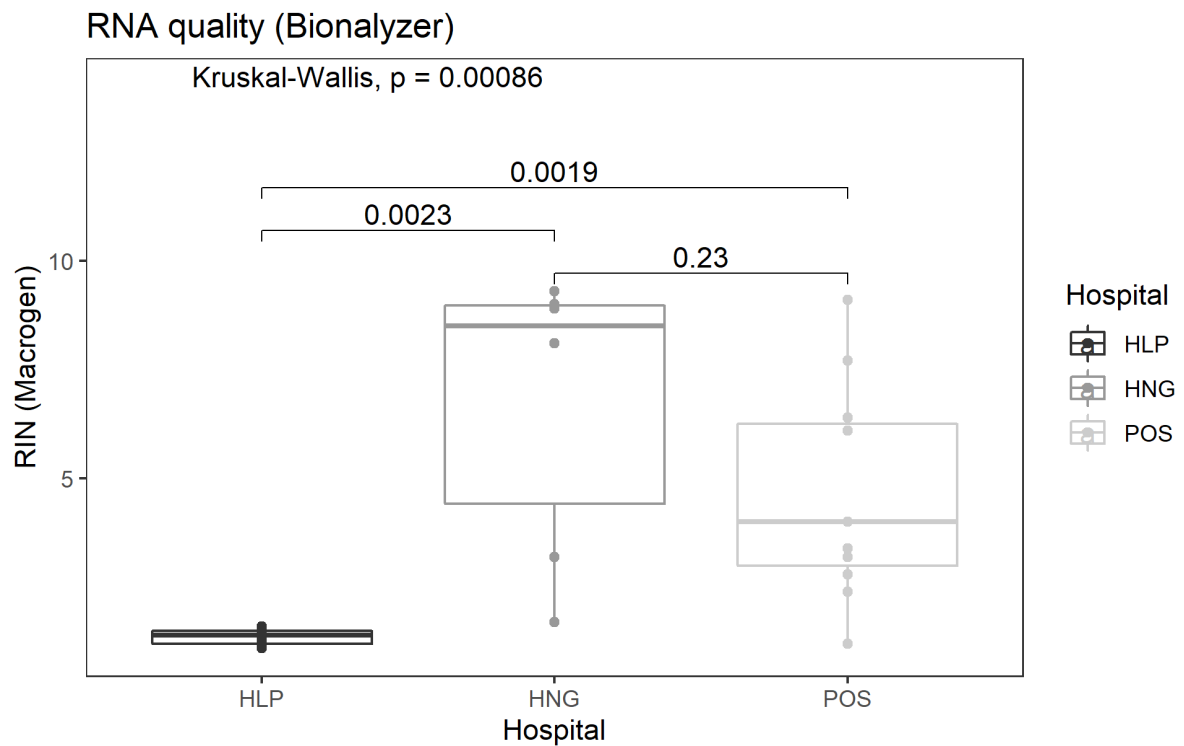
Graphics. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 3, Issue 2, pp. 180–185). <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>

Witkowski, M. T., Dolgalev, I., Evensen, N. A., Ma, C., Chambers, T., Roberts, K. G., Sreeram, S., Dai, Y., Tikhonova, A. N., Lasry, A., Qu, C., Pei, D., Cheng, C., Robbins, G. A., Pierro, J., Selvaraj, S., Mezzano, V., Daves, M., Lupo, P. J., ... Aifantis, I. (2020). Extensive Remodeling of the Immune Microenvironment in B Cell Acute Lymphoblastic Leukemia. *Cancer Cell*, 37(6), 867-882.e12. <https://doi.org/10.1016/j.ccell.2020.04.015>

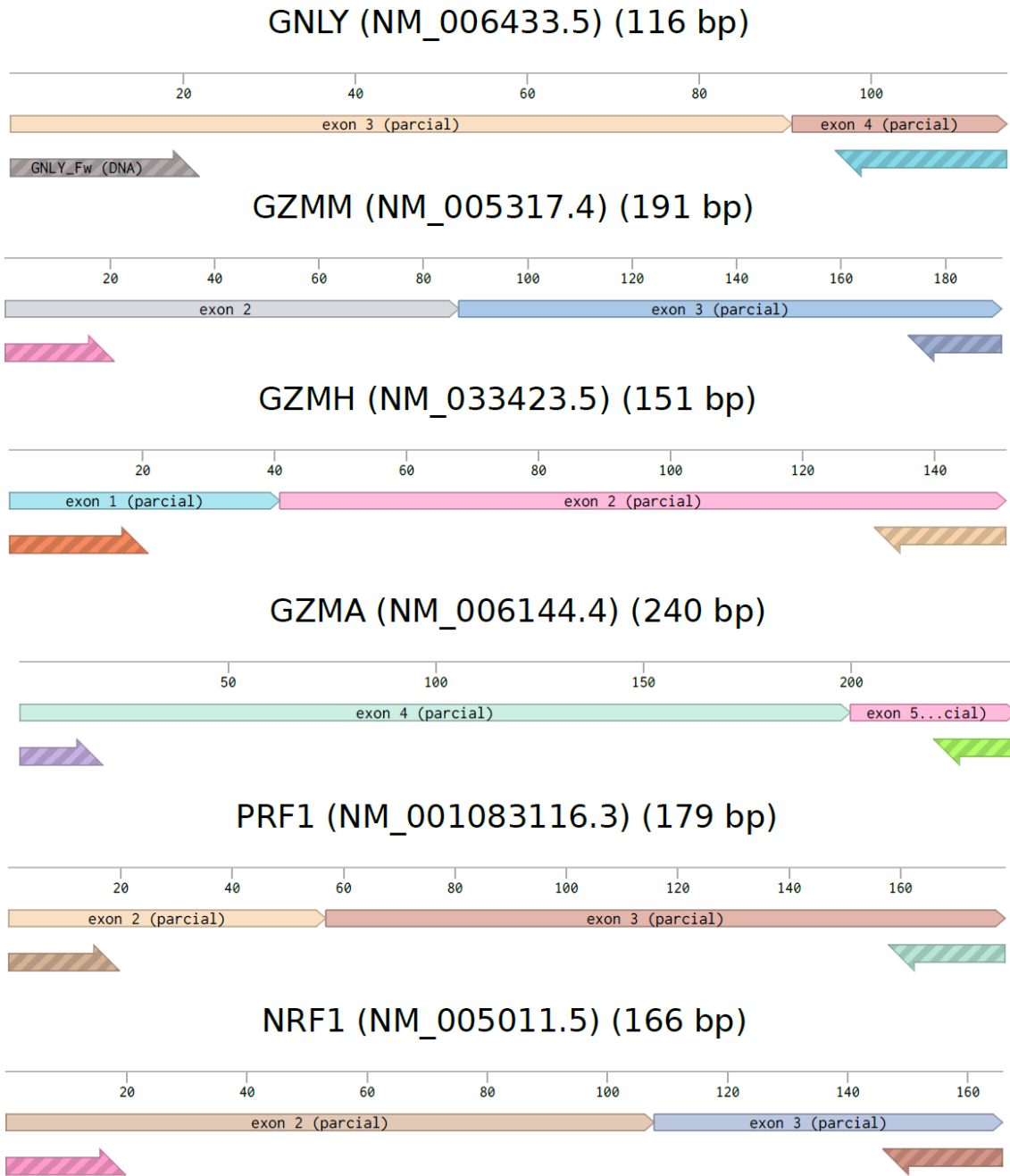
Xie, Y., Akpınarli, A., Maris, C., Hipkiss, E. L., Lane, M., Kwon, E. K. M., Muranski, P., Restifo, N. P., & Antony, P. A. (2010). Naive tumor-specific CD4+ T cells differentiated in vivo eradicate established melanoma. *Journal of Experimental Medicine*, 207(3), 651–667. <https://doi.org/10.1084/jem.20091921>

Zhang, R., Zhu, H., Yuan, Y., Zhao, J., Yang, X., & Tian, Z. (2020). Risk factors for relapse of childhood b cell acute lymphoblastic leukemia. *Medical Science Monitor*, 26, 1–7. <https://doi.org/10.12659/MSM.923271>

## VII) Anexo



**Figura S.1:** Calidad de las muestras de RNA secuenciadas por RNA-seq y utilizadas en este estudio, clasificadas según su hospital de origen. El RIN es un parámetro indicador de la calidad del RNA, que va de 0 a 10, siendo 10 la máxima calidad.



**Figura S.2:** Esquema del sitio de pegado de *primers* utilizados para el cálculo del *cytolytic score*.



**Tabla S.1: Resultados de análisis estadístico entre proporciones de poblaciones celulares y clusters determinados por heatmap ajustando datos a una distribución beta con inflados en cero.**

<b>TIL10</b>				
<b>Población</b>	<b>Supuestos</b>	<b>p-val Cluster 1-2</b>	<b>p-val Cluster 1-3</b>	<b>p-val Cluster 2-3</b>
<i>B cells</i>	Se cumplen	0,335	$1,87 \times 10^{-4}$	$1,35 \times 10^{-5}$
<i>Monocytes</i>	No se cumplen	----	----	----
<i>Neutrophils</i>	Se cumplen	0,160	0,495	0,0223
<i>NK cells</i>	Se cumplen	0,576	0,277	0,786
<i>T cells CD4</i>	Se cumplen	0,533	$2,99 \times 10^{-5}$	$1,3 \times 10^{-5}$
<i>T cells CD8</i>	Se cumplen	0,771	0,231	0,424
<i>Dendritic cells</i>	Se cumplen	0,895	0,852	0,795
<b>LM22</b>				
<b>Población</b>	<b>Supuestos</b>	<b>p-val Cluster 1-2</b>	<b>p-val Cluster 1-3</b>	<b>p-val Cluster 2-3</b>
<i>B cells naive</i>	Se cumplen	0,560	0,548	0,219
<i>Neutrophils</i>	Se cumplen	0,455	0,569	
<i>Eosinophils</i>	Se cumplen	0,173	0,073	0,562
<i>Follicular Helper</i>	Se cumplen	0,531	0,584	0,918
<i>Monocytes</i>	No se cumplen	----	----	----
<i>NK resting</i>	Se cumplen	0,466	0,677	0,768
<i>T cells CD4<sup>+</sup> memory resting</i>	No se cumplen	----	----	----
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,80223	$6,55 \times 10^{-3}$	0,01323
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,272	0,171	0,965
<i>Plasma cells</i>	Se cumplen	0,139	0,0214	0,5002

**Tabla S.1:** El modelo estadístico se realizó con la función `glmmTMB`. Se utilizaron las proporciones obtenidas a partir de MIXTURE utilizando los datos de RNA-seq, TPM habiendo sido corregidos por la función `ComBat` que fueron las mismas utilizadas para realizar el heatmap a partir del cual se agruparon a los pacientes en clusters. Solo se analizaron aquellas poblaciones que tuviesen al menos más de 3 muestras con proporciones distintas de cero. Las células de linaje B no fueron analizadas.

**Tabla S.2: Resultados de prueba de Chi cuadrado de *Pearson* entre variables clínicas y clusters determinados por el agrupamiento no supervisado.**

<b>clusters TIL10</b>		
<b>Variable clínica</b>	<b>Chi-cuadrado</b>	<b>p-valor</b>
Muerte y/o recaída	2,7391	0,2542
Toxicidad aguda relacionada al tratamiento	1,9886	0,37
Respuesta a prednisona	3,0435	0,2183
Grupo de riesgo	7,08	0,1317
EMRCat	7,3805	0,1171
<b>clusters LM22</b>		
<b>Variable clínica</b>	<b>Chi-cuadrado</b>	<b>p-valor</b>
Muerte y/o recaída	1,5606	0,4583
Toxicidad aguda relacionada al tratamiento	1,9071	0,9091
Respuesta a prednisona	3,5583	0,1688
Grupo de riesgo	5,0551	0,2817
EMRCat	7,8813	0,0960

**Tabla S.2:** Se utilizó la función *chisq.test* en R para realizar el análisis.

**Tabla S.3: Análisis estadístico entre proporciones de poblaciones celulares y variables clínicas ajustando datos a una distribución beta con inflados en cero**

<b>Variable clínica: Muerte y/o recaída</b>		
<b>Población celular (LM22)</b>	<b>Supuestos</b>	<b>p- valor</b>
<i>Neutrophils</i>	Se cumplen	0,548
<i>Eosinophils</i>	Se cumplen	0,828
<i>Macrophages M0</i>	Se cumplen	----
<i>Monocytes</i>	Se cumplen	0,878
<i>NK resting</i>	Se cumplen	0,316
<i>T cells CD4<sup>+</sup> memory resting</i>	Se cumplen	0,255
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,808
<i>T cells CD8<sup>+</sup></i>	No se cumplen	----

<b>Variable clínica: Toxicidad aguda relacionada al tratamiento</b>		
<b>Población celular (LM22)</b>	<b>Supuestos</b>	<b>p- valor</b>
<i>Neutrophils</i>	Se cumplen	0,554
<i>Eosinophils</i>	Se cumplen	0,726
<i>Macrophages M0</i>	Se cumplen	----
<i>Monocytes</i>	Se cumplen	0,225
<i>NK resting</i>	Se cumplen	0,00854
<i>T cells CD4<sup>+</sup> memory resting</i>	Se cumplen	0,464
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,0801
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,0438
<b>Variable clínica: Respuesta a prednisona</b>		
<b>Población celular (LM22)</b>	<b>Supuestos</b>	<b>p- valor</b>
<i>Neutrophils</i>	Se cumplen	0,633
<i>Eosinophils</i>	Se cumplen	0,887
<i>Macrophages M0</i>	Se cumplen	0,606
<i>Monocytes</i>	Se cumplen	0,139
<i>NK resting</i>	Se cumplen	0,762
<i>T cells CD4<sup>+</sup> memory resting</i>	Se cumplen	0,0527
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,86
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,924
<b>Variable clínica: EMRCat</b>		
<b>Población celular (LM22)</b>	<b>Supuestos</b>	<b>p- valor</b>
<i>Neutrophils</i>	Se cumplen	0,573
<i>Eosinophils</i>	Se cumplen	0,0904
<i>Macrophages M0</i>	Se cumplen	0,93
<i>Monocytes</i>	Se cumplen	0,272
<i>NK resting</i>	No se cumplen	0,207
<i>T cells CD4<sup>+</sup> memory resting</i>	Se cumplen	0,594
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,503

<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,924		
<b>Variable clínica: Grupo de riesgo</b>				
<b>Población celular (LM22)</b>	<b>Supuestos</b>	<b>p-valor (Estándar - intermedio)</b>	<b>p-valor (Estándar - alto)</b>	<b>p-valor (Intermedio - alto)</b>
<i>Neutrophils</i>	No se cumplen	----	----	----
<i>Eosinophils</i>	No se cumplen	----	----	----
<i>Macrophages M0</i>	No se cumplen	----	----	----
<i>Monocytes</i>	Se cumplen	0,4951	0,7703	0,9062
<i>NK resting</i>	Se cumplen	0,8971	0,8251	0,9587
<i>T cells CD4<sup>+</sup> memory resting</i>	Se cumplen	0,9288	0,9245	0,9987
<i>T cells CD4<sup>+</sup> naive</i>	Se cumplen	0,98	0,9329	0,94
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,7829	0,9969	0,8309
<b>Variable clínica: Muerte y/o recaída</b>				
<b>Población celular (TIL10)</b>	<b>Supuestos</b>	<b>p- valor</b>		
<i>T cells CD4<sup>+</sup></i>	Se cumplen	0,0405		
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,193		
<i>Dendritic cells</i>	Se cumplen	0,882		
<i>NK cells</i>	Se cumplen	0,558		
<i>Monocytes</i>	Se cumplen	0,0234		
<i>Neutrophils</i>	No se cumplen	----		
<b>Variable clínica: Toxicidad aguda relacionada al tratamiento</b>				
<b>Población celular (TIL10)</b>	<b>Supuestos</b>	<b>p- valor</b>		
<i>T cells CD4<sup>+</sup></i>	Se cumplen	0,356		
<i>T cells CD8<sup>+</sup></i>	Se cumplen	1,00x10 <sup>-6</sup>		
<i>Dendritic cells</i>	Se cumplen	0,58		
<i>NK cells</i>	Se cumplen	0,73		
<i>Monocytes</i>	Se cumplen	0,437		
<i>Neutrophils</i>	Se cumplen	0,288		

<b>Variable clínica: Respuesta a prednisona</b>				
<b>Población celular (TIL10)</b>	<b>Supuestos</b>	<b>p- valor</b>		
<i>T cells CD4<sup>+</sup></i>	No se cumplen	----		
<i>T cells CD8<sup>+</sup></i>	Se cumplen	0,494		
<i>Dendritic cells</i>	Se cumplen	0,614		
<i>NK cells</i>	Se cumplen	0,743		
<i>Monocytes</i>	Se cumplen	0,974		
<i>Neutrophils</i>	Se cumplen	0,906		
<b>Variable clínica: EMRCat</b>				
<b>Población celular (TIL10)</b>	<b>Supuestos</b>	<b>p- valor</b>		
<i>T cells CD4<sup>+</sup></i>	No se cumplen	----		
<i>T cells CD8<sup>+</sup></i>	No se cumplen	----		
<i>Dendritic cells</i>	Se cumplen	0,475		
<i>NK cells</i>	Se cumplen	0,394		
<i>Monocytes</i>	No se cumplen	----		
<i>Neutrophils</i>	No se cumplen	----		
<b>Variable clínica: Grupo de riesgo</b>				
<b>Población celular (TIL10)</b>	<b>Supuestos</b>	<b>p-valor (Estándar - intermedio)</b>	<b>p-valor (Estándar - alto)</b>	<b>p-valor (Intermedio - alto)</b>
<i>T cells CD4<sup>+</sup></i>	Se cumplen	0,4906	0,893	0,8242
<i>T cells CD8<sup>+</sup></i>	No se cumplen	----	----	----
<i>Dendritic cells</i>	No se cumplen	----	----	----
<i>NK cells</i>	No se cumplen	0,8024	0,5149	0,6791
<i>Monocytes</i>	No se cumplen	----	----	----
<i>Neutrophils</i>	Se cumplen	0,916	0,9786	0,9573

**Tabla S.3:** El modelo estadístico se realizó con la función `glmmTMB`. Se utilizaron las proporciones obtenidas a partir de MIXTURE utilizando los datos de RNA-seq, TPM. Solo se analizaron aquellas poblaciones que tuviesen al menos más de 3 muestras con proporciones distintas de cero. Las células de linaje B no fueron analizadas. Se incluyó al lote de secuenciación como variable aleatoria.

**Tabla S.4: Resultados de curvas de Kaplan Meier dividiendo a los pacientes en cuanto a la proporción de las distintas poblaciones celulares estimadas por MIXTURE y utilizando distintas variables como evento para medir la sobrevida libre de evento.**

<b>LM22</b>									
	<b>Sobrevida global</b>			<b>Sobrevida libre de recaída</b>			<b>Sobrevida libre de evento</b>		
<b>Población</b>	HR	CoxP	Log-rank	HR	CoxP	Log-rank	HR	CoxP	Log-rank
Linfocitos T CD4 <sup>+</sup> memory resting	4,244	<b>0,085</b>	0,114	3,124	0,218	0,236	2,338	0,239	0,25
Linfocitos T CD8 <sup>+</sup>	3,687	0,126	0,153	0,437	0,449	0,462	1,716	0,454	0,46
Neutrófilos	2,081	0,412	0,422	0,579	0,629	0,633	0,873	0,87	0,87
Eosinófilos	0,472	0,493	0,503	0,292	0,242	0,272	0,679	0,594	0,596
Células NK	0,418	0,324	0,339	3,024	0,304	0,327	1,1374	0,661	0,662
Monocitos	0,381	0,272	0,291	0,391	0,293	0,31	0,631	0,516	0,519
Linfocitos T CD4 <sup>+</sup> naive	0,263	0,199	0,232	2,216	0,38	0,392	0,516	0,412	0,42
Macrófagos M0	0	0,375	0,999	0	0,304	0,999	0	0,215	0,999
Linfocitos T CD8 <sup>+</sup> + NK reposo + NK activadas	ND	ND	ND	2,563	0,399	0,416	5,393	<b>0,081</b>	0,119
<b>TIL10</b>									
	<b>Sobrevida global</b>			<b>Sobrevida libre de recaída</b>			<b>Sobrevida libre de evento</b>		
<b>Población</b>	HR	CoxP	Log-rank	HR	CoxP	Log-rank	HR	CoxP	Log-rank
Neutrófilos	3,238	0,173	0,198	0,437	0,395	0,409	0,519	0,414	0,422
Linfocitos T CD8 <sup>+</sup>	2,347	0,336	0,351	3,544	0,177	0,206	3,521	0,101	0,124
Monocitos	2,008	0,436	0,445	1,569	0,628	0,63	1,969	0,376	0,384

Células Dendríticas	1,766	0,528	0,533	1,883	0,484	0,491	2,841	0,192	0,212
Linfocitos T CD4 <sup>+</sup>	1,637	0,586	0,59	2,22	0,373	0,385	1,609	0,535	0,539
células NK	0,387	0,379	0,396	ND	ND	ND	2,386	0,29	0,304

**Tabla S.4:** Valores de HR, p-valor de la prueba de Log-rank y p-valor del HR (Cox P) correspondientes a cada población celular en las curvas de supervivencia global, libre de recaída o libre de evento (muerte y/o recaída). ND (no determinado): al momento de establecer el punto de corte para separar a las muestras en 2 grupos, alguno de los dos no poseía evento, con lo cual el valor resultante carecía de sentido.

**Tabla S.5: Resultados de prueba de Wilcoxon entre proporciones de poblaciones celulares estimadas por MIXTURE (TIL10 y LM22) de pacientes de LLA-B y donantes sanos**

<b>TIL10</b>	
<b>Población</b>	<b>p-valor</b>
<i>B cells</i>	5,52x10 <sup>-8</sup>
<i>Monocytes</i>	9,73x10 <sup>-7</sup>
<i>T cells CD8<sup>+</sup></i>	1,09x10 <sup>-6</sup>
<i>Tregs</i>	2,43x10 <sup>-5</sup>
<i>T cells CD4<sup>+</sup></i>	2,62x10 <sup>-5</sup>
<i>Neutrophils</i>	0,1614
<i>NK cells</i>	0,3642
<i>Dendritic cells</i>	0,4624
<i>Macrophages M0</i>	0,6757
<b>LM22</b>	
<b>Población</b>	<b>p-valor</b>
<i>B cells naive</i>	1,32x10 <sup>-6</sup>
<i>Monocytes</i>	1,33x10 <sup>-5</sup>
<i>T cells CD8<sup>+</sup></i>	5,54x10 <sup>-5</sup>
<i>T cells CD4<sup>+</sup> memory resting</i>	6,04x10 <sup>-4</sup>
<i>Eosinophils</i>	7,03x10 <sup>-3</sup>
<i>NK resting</i>	0,0188

<i>Neutrophils</i>	0,0766
<i>Plasma cells</i>	0,1003
<i>Macrophages M0</i>	0,2743
<i>B cells memory</i>	0,5213
<i>NK activated</i>	0,5213
<i>T cells CD4<sup>+</sup> naive</i>	0,6225
<i>T cells follicular helpers</i>	0,6757
<i>Macrophages M2</i>	0,6757

**Tabla S.5:** P valores de la prueba de Wilcoxon para comparar proporciones de poblaciones celulares estimadas por MIXTURE (TIL10 y LM22) entre pacientes de LLA-B y donantes sanos. Solo se informan las poblaciones celulares para las cuales se pudo realizar la prueba de Wilcoxon.