



# UNIVERSIDAD DE GRANADA

Programa de Doctorado en Bioquímica y Biología Molecular (B16.56.1)

————— Tesis Doctoral —————

## **Análisis de la implicación de los ARN largos no codificantes y el complejo SWI/SNF en cáncer**

**Daniel Jesús García García**

**Directores de Tesis:**

Dr. Pedro Pablo Medina Vico

Dr. Marta Eugenia Cuadros Celorrio

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Daniel Jesús García García  
ISBN: 978-84-1195-492-1  
URI: <https://hdl.handle.net/10481/96707>



# Índice

Resumen.....	14
Objetivos .....	17
CAPÍTULO I: Introducción .....	19
1.1 Cáncer: Una enfermedad mundial.....	19
1.1.1 Estadísticas globales del cáncer .....	19
1.1.2 Cáncer infantil: incidencia, mortalidad y subtipos .....	20
1.2 Bases moleculares del desarrollo tumoral .....	21
1.2.1 Principios de la evolución tumoral.....	21
1.2.1 Bases genéticas de la tumorigénesis.....	22
1.2 Leucemia Linfoblástica Aguda .....	23
1.2.1 Definición, tipos e incidencia mundial.....	23
1.2.2 Etiología.....	24
1.2.2.3 Modelo del doble evento .....	25
1.2.3 Evaluación y diagnóstico .....	26
1.2.4 Predisposición genética .....	27
1.2.5 Tratamientos.....	32
1.3 Regulación de la expresión génica .....	33
1.3.1 Epigenética.....	33
1.3.2 Mecanismos de regulación de la expresión génica.....	34
1.4 Complejo regulador de la cromatina SWI/SNF .....	36
1.4.1. cBAF: Las subunidades ARID1 .....	38
1.4.2 PBAF: La subunidad ARID2 .....	41
1.5 ARN no codificantes .....	42
1.5.1 Definición de ARNlnc.....	43
1.5.2 Síntesis, procesamiento y localización de los ARNlnc .....	43
1.5.3 Clasificación de los ARNlnc .....	44
1.5.4 Funciones biológicas de los ARNlnc .....	45
CAPÍTULO II: Materiales y Métodos .....	50
2.1 Análisis estadístico.....	50

2.2 Programas, herramientas online y bases de datos .....	50
2.3 Microarray de ARNlnc .....	51
2.3.1 Muestras de pacientes con LLA-B y sanas.....	51
2.3.2 Diseño del microarray y predicción de la regulación de ARNlnc .....	52
2.3.3 Etiquetado de ARN e hibridación.....	52
2.3.4 Análisis de datos de expresión génica .....	53
2.3.5 Análisis de ontología de genes .....	53
2.4 Datos externos: TCGA y CCLE.....	54
2.4.1 Análisis de datos de expresión de pacientes.....	54
2.4.2 Análisis de datos de expresión de líneas celulares .....	56
2.5 Secuenciación ARN-Seq .....	57
2.5.1 Genotipo de ratones.....	57
2.5.2 Muestras .....	57
2.5.3 Protocolo de secuenciación .....	57
2.6. Métodos de biología molecular .....	59
2.6.1 Extracción y análisis de ARNm .....	59
2.6.2 RT-qPCR.....	60
2.6.3 Extracción, separación y cuantificación de proteínas.....	60
2.6.4 Cultivos celulares .....	62
2.6.5 Silenciamiento génico mediante ARNsi.....	63
2.6.6 Ensayos de viabilidad celular .....	63
2.6.7 Ensayos de proliferación celular .....	64
2.6.8 Evaluación de parámetros biológicos de ratones.....	65
CAPÍTULO III: ARNlnc en LLA pediátrica.....	67
3.1 Antecedentes .....	67
3.1.1 ARNlnc reguladores de la hematopoyesis.....	67
3.1.2 ARNlnc TCL6: Biomarcador en LLA-B pediátrica .....	68
3.2 Resultados .....	70
3.2.1 Firma de expresión génica aberrante en LLA-B pediátrica.....	70
3.2.2 Rutas biológicas asociadas al perfil de ARNlnc diferencialmente expresados en LLA-B.....	71

3.2.3 Altos niveles de expresión génica de la pareja <i>AL133346.1/CCN2</i> en LLA-B .....	73
3.2.4 Expresión de <i>AL133346.1/CCN2</i> correlacionada en LLA-B .....	75
3.2.5 Análisis <i>in silico</i> de los mecanismos de regulación de <i>AL133346.1/CCN2</i> .....	76
3.2.6 Altos niveles de expresión de <i>CCN2</i> asociados con un mejor pronóstico en pacientes con LLA-B .....	77
4. Discusión .....	80
CAPÍTULO IV: Implicación de <i>ARID2</i> en LLA .....	85
4.1 Antecedentes .....	85
4.1.1 Funciones y alteraciones del complejo SWI/SNF en la hematopoyesis .....	85
4.1.2 El papel de <i>ARID2</i> en la hematopoyesis .....	86
4.2 Resultados .....	88
4.2.1 Valor pronóstico de <i>ARID2</i> en pacientes pediátricos con LLA-B .....	88
4.2.2 Niveles de expresión génica de <i>ARID2</i> en líneas celulares de LLA pediátricas .....	90
4.2.3 Optimización del silenciamiento génico de <i>ARID2</i> en líneas celulares de LLA-B .....	91
4.2.4 El silenciamiento de <i>ARID2</i> disminuye la viabilidad y proliferación celular .....	93
4.2.5 Menor expresión génica de <i>ARID2</i> en pacientes pediátricos sanos .....	95
4.3 Discusión .....	97
CAPÍTULO V: Implicación de <i>ARID1A</i> y <i>ARID1B</i> en cáncer .....	102
5.1 Antecedentes .....	102
5.1.1 Complejos SWI/SNF residuales en el desarrollo tumoral .....	102
5.1.2 Letalidad sintética de subunidades del complejo SWI/SNF .....	103
5.2 Resultados .....	104
5.2.1 Parámetros biológicos afectados en los ratones DKO <i>Arid1a/Arid1b</i> .....	104
5.2.2 Análisis histológico de los ratones DKO <i>Arid1a/Arid1b</i> .....	105
5.2.3 Menor expresión de marcadores de células madre intestinales en el DKO <i>Arid1a/Arid1b</i> .....	107
5.2.4 Control de calidad de los datos de secuenciación .....	108

5.2.5 Selección de las muestras para el análisis de expresión diferencial .	110
5.2.6 Análisis de expresión diferencial y GSEA de MEF .....	112
5.2.7 Análisis de expresión diferencial y GSEA de los tejidos de hígado y yeyuno .....	113
5.3 Discusión.....	118
Conclusiones .....	124
Material Suplementario .....	127
Bibliografía.....	137
Anexo .....	169
Financiación .....	169
Publicaciones.....	169
Permisos de derecho de autor .....	171

# Resumen

# Resumen

A pesar de los avances médicos que se han producido en las últimas décadas en cuanto a prevención, diagnóstico y terapia, la incidencia del cáncer y sus tasas de mortalidad continúan siendo elevadas a nivel mundial. En la actualidad, el campo de la genómica y la bioinformática están experimentando un período de marcado crecimiento en la investigación de esta patología. El objetivo fundamental que persiguen estas disciplinas consiste en el esclarecimiento de los mecanismos moleculares subyacentes del desarrollo tumoral, así como la aplicación práctica de tales descubrimientos en el ámbito clínico, con el propósito de mejorar el proceso de diagnóstico, el pronóstico y el tratamiento del cáncer. Específicamente, el cáncer es la principal causa de muerte infantil relacionada con enfermedades, siendo la leucemia linfoblástica aguda la neoplasia hematológica con mayor incidencia. Aunque se ha alcanzado una tasa de cura elevada con la mejora de la quimioterapia, aquellos pacientes pediátricos que no responden al tratamiento aún siguen representando una de las principales causas de muerte infantil.

Hasta ahora, numerosos estudios han identificado alteraciones genéticas en genes codificantes de proteínas que han permitido clasificar molecularmente el cáncer. Sin embargo, el estudio de los genes no codificantes de proteínas y la investigación sobre las alteraciones epigenéticas en cáncer, que no implican cambios en la secuencia del ADN, han experimentado un notable avance en los últimos años que ha fundamentado su vinculación con esta enfermedad.

Aunque la gran mayoría de estudios se han centrado preferentemente en genes codificantes de proteínas, existe otro tipo de moléculas reguladoras del control de la expresión génica denominadas ARN largos no codificantes, que aparecen alteradas en un gran número de enfermedades. A pesar de la comprensión actual de las bases genéticas de la leucemia linfoblástica aguda, la detección de alteraciones en los ARN largos no codificantes podría mejorar el diagnóstico de los pacientes y estratificarlos de una manera más precisa para el tratamiento. Así, en esta tesis, se ha descubierto que los niveles del ARNlnc *AL133346.1* y su gen codificante de proteína solapante *CCN2* se encontraban sobre-expresados transcripcionalmente en la leucemia linfoblástica aguda de subtipo celular B en comparación con muestras sanas y muestras de leucemia linfoblástica aguda de subtipo celular T, por lo que podrían ser utilizados como biomarcadores para el diagnóstico de esta enfermedad. Aunque no se pudo demostrar a nivel mecanístico, se hipotetizó que la sobre-expresión correlacionada de ambos genes podría estar justificada por la presencia de elementos reguladores comunes o por la ac-

ción reguladora en *cis* que ejerce el ARNlnc *AL133346.1* sobre *CCN2*. Desde un punto de vista relevante para la clínica, en relación con el pronóstico de los pacientes con leucemia linfoblástica aguda pediátricos, se ha demostrado que los niveles de sobre-expresión de *CCN2* podría considerarse un factor pronóstico independiente que mejora la supervivencia global de los pacientes (Cociente de Riesgo: 0,566; Intervalo Confianza del 95%: 0,328-0,980).

En otro contexto, en relación con la epigenética, cabe destacar el control regulatorio de la expresión génica ejercido por el complejo remodelador de la cromatina SWI/SNF. Los avances en las tecnologías de secuenciación han permitido identificar mutaciones en las subunidades que componen el complejo SWI/SNF en hasta el 25% de los casos de cáncer, lo que pone de manifiesto su importancia en la carcinogénesis. En esta línea, se ha analizado los niveles de transcripción de las distintas subunidades y su correlación con la supervivencia en pacientes pediátricos con leucemia linfoblástica aguda de células B. *ARID2*, reconocido como un gen supresor tumoral, fue la que mostró el impacto más significativo en la supervivencia de los pacientes, asociándose sus mayores niveles de expresión con tiempos de supervivencia global más largos. Por otro lado, *ARID2* se encontraba sobre-expresado en comparación con muestras sanas, lo que sugería que podía poseer un papel dual como supresor tumoral y factor en el mantenimiento tumoral. Esta hipótesis fue evaluada en experimentos *in vitro*, donde la inhibición de *ARID2* mediante ARN de silenciamiento redujo la viabilidad celular de líneas en las líneas de LLA-B RS4;11 y SUP-B15.

Por último, con el fin de caracterizar cómo los complejos SWI/SNF defectuosos contribuyen a la carcinogénesis, se evaluó el impacto biológico que podría producir la delección de las subunidades parálogas *Arida* y *Arid1b* del subtipo de complejo SWI/SNF cBAF en modelos animales. La eliminación de ambos genes produjo una mortalidad acelerada de estos ratones doble *knockout* en comparación con los silvestres. A nivel tisular, los ratones doble *knockout* mostraron signos patológicos tanto en el hígado como en las criptas del intestino delgado. La secuenciación ARN-Seq de muestras de estos tejidos mostró un gran número de genes expresados diferencialmente que enriquecían rutas biológicas relacionadas con la apoptosis, la alteración del metabolismo de los nutrientes y la señalización de la inflamación. En futuros experimento, se pretende estudiar los elementos reguladores controlados de manera específica por estas subunidades con el fin de conocer las rutas biológicas que controla de manera directa el complejo cBAF.

# Objetivos

# Objetivos

El objetivo general de esta tesis trata de profundizar en el estudio de los genes no codificantes de proteínas y los complejos remodeladores de la cromatina para identificar nuevos biomarcadores de utilidad en oncología clínica. Para ello, se plantearon los siguientes objetivos específicos:

**Objetivo 1.** Determinación de ARN largos no codificantes alterados en pacientes pediátricos con leucemia linfoblástica aguda de células B con el fin de determinar nuevos biomarcadores en esta enfermedad.

**Objetivo 2.** Análisis de las alteraciones genéticas de las subunidades del complejo SWI/SNF en cáncer:

**2.1** Estudiar las alteraciones genéticas de ARID2 en leucemia linfoblástica aguda con el propósito de analizar sus efectos y comprender mejor el papel de esta subunidad en el desarrollo de las neoplasias hematológicas.

**2.2** Estudiar el fenotipo que produce eliminar los genes *Arid1a* y *Arid1b* en modelos *in vivo* para comprender el impacto y el papel biológico de estas subunidades en el desarrollo tumoral.

# Capítulo I

# Introducción

# CAPÍTULO I: Introducción

## 1.1 Cáncer: Una enfermedad mundial

El cáncer es considerado una enfermedad genética caracterizada por un crecimiento celular descontrolado que puede conducir a la propagación de células anormales a otras partes del cuerpo mediante invasión local y/o metástasis (Maru, 2016). Las alteraciones genéticas originarias de esta enfermedad pueden ser heredadas, ocasionadas por errores durante la división celular o por la influencia de factores ambientales externos, tales como el consumo de tabaco o el tipo de dieta (Jassim et al., 2023). La identificación de estas alteraciones genéticas podría revelar nuevas vías terapéuticas para avanzar en la investigación de esta enfermedad compleja.

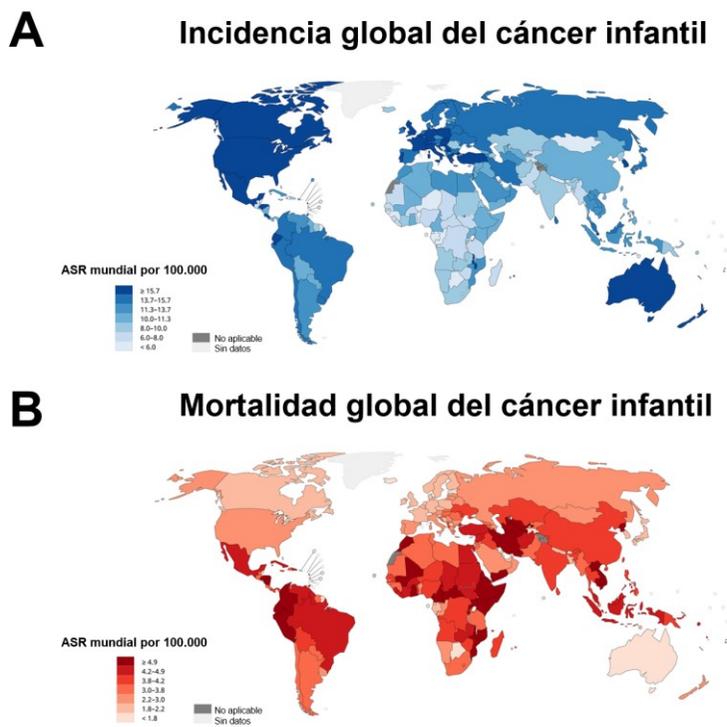
### 1.1.1 Estadísticas globales del cáncer

El cáncer es una enfermedad multifactorial que afecta principalmente a personas mayores de 50 años. A pesar de los avances en la prevención, el diagnóstico y el tratamiento del cáncer, de acuerdo con las estadísticas recogidas por la Organización Mundial de la Salud en 2019, se trata de la primera o segunda principal causa de muerte antes de los 70 años en más del 60% de los países estudiados. El envejecimiento y el crecimiento de la población reflejan el aumento de la incidencia y de la mortalidad del cáncer en todo el mundo (Sung et al., 2021).

De acuerdo con el informe estadístico GLOBOCAN 2020, que describe la magnitud y la distribución de los principales tipos de cáncer en distintas regiones mundiales, se estimaron 19,3 millones de nuevos casos y 10 millones de muertes debido al cáncer cada año. El cáncer de mama es el tipo más comúnmente diagnosticado (11,7% de los casos totales) y principal causa de muerte en mujeres, mientras que el cáncer de pulmón representa la causa de muerte más frecuente (18,0% de las muertes por cáncer totales), siendo la dominante en hombres. A nivel mundial, se estima que para el año 2.040 se producirá un aumento del 47% en la incidencia de cáncer hasta 28,4 millones de casos nuevos. Además del envejecimiento de la población, existe un aumento de la prevalencia de otros factores de riesgo como el tabaquismo, la inactividad física o el aumento de peso corporal, principalmente en países con un Índice de Desarrollo Humano (IDH) bajo o medio, que podría provocar en las próximas décadas un cambio en los perfiles de tipos de cáncer más prevalentes (Sung et al., 2021).

### 1.1.2 Cáncer infantil: incidencia, mortalidad y subtipos

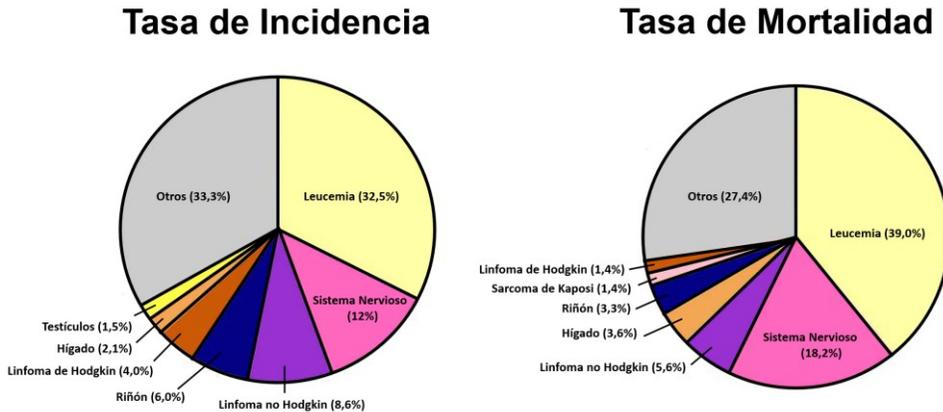
El cáncer es la principal causa de muerte infantil relacionada con enfermedades (Cunningham et al., 2018; Huang et al., 2023). En 2020, se diagnosticaron hasta 207.000 nuevos casos de cáncer infantil alrededor del mundo. Las tasas más altas de incidencia fueron significativamente mayores en países con un alto IDH, sobre todo, en América del Norte y en el occidente europeo [Figura 1A]. En el mismo año, se llegaron a alcanzar hasta las 80.000 muertes debido a esta enfermedad, siendo mayor la frecuencia en regiones con menor IDH, tales como el este africano o el sureste asiático [Figura 1B] (Huang et al., 2023). No obstante, en las últimas décadas se ha conseguido reducir significativamente la mortalidad del cáncer infantil gracias a los avances quimioterapéuticos (van Kalsbeek et al., 2023).



**Figura 1.** Estimación de la incidencia (A) y la mortalidad (B) mundial del cáncer infantil (0-14 años) para ambos sexos en 2020. ASR (del inglés, “Age-Standardised Rate”) representa la tasa estandarizada por edad. Adaptación de (Huang et al., 2023).

Entre los distintos tipos de cáncer infantiles, durante el año 2018 la frecuencia de los casos diagnosticados fue mayor para la leucemia (32,5%), seguido

de tumores relacionados con el sistema nervioso (12%), linfomas no Hodgkin (8,6%) y cáncer renal (6,0%). Para pacientes pediátricos menores de 14 años, la leucemia resultó ser además la principal causa de muerte infantil (39,0%) en comparación con los diferentes tipos de cáncer [Figura 2] (Namayandeh et al., 2020).



**Figura 2.** A la izquierda, tasas de incidencia (N = 200.166 casos) y, a la derecha, tasas de mortalidad (N = 74.956 muertes) de los distintos subtipos de cáncer pediátrico en pacientes menores de 19 años para ambos sexos en 2018. Adaptación de (Namayandeh et al., 2020).

## 1.2 Bases moleculares del desarrollo tumoral

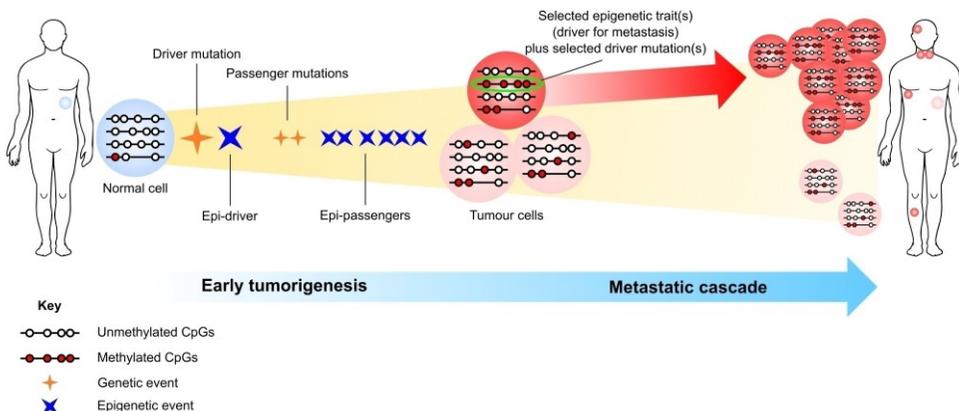
### 1.2.1 Principios de la evolución tumoral

Los modelos actuales que explican la dinámica del desarrollo tumoral se encuentran basados en una analogía con los principios de la evolución darwiniana. Según estos modelos, la mayoría de los tumores se originan a partir de una única célula neoplásica que evoluciona a través de un proceso de selección de alteraciones somáticas, produciendo la proliferación y la supervivencia de los clones más aventajados derivados de esta célula inicial. Por ello, se considera que los tumores están compuestos por múltiples subclones originados a partir de un ancestro común y que se expanden simultáneamente (Nicoś & Krawczyk, 2022; Vendramin et al., 2021). Cuando dos subclones crecen independientemente, cada uno de ellos podría adquirir nuevas ventajas selectivas distintas del otro. Los patrones genéticos de heterogeneidad intratumoral y la diversidad subclonal son consecuencia directa de esta dinámica evolutiva del desarrollo tumoral (West et al., 2021).

## 1.2.1 Bases genéticas de la tumorigénesis

El origen del cáncer se basa en la acumulación de alteraciones genéticas y epigenéticas, que aumentan la inestabilidad genómica celular. Esta desestabilización del genoma de las células tumorales promueve su modificación constante a través de la adquisición de nuevas alteraciones en el ADN, lo que conlleva a la evolución continua de las subpoblaciones clonales y al aumento de heterogeneidad en el tumor (Pikor et al., 2013).

Las células transformadas genéticamente se caracterizan por su mayor capacidad de proliferación, su resistencia a la muerte celular, su habilidad para inducir la angiogénesis y su capacidad para invadir otros tejidos (Hanahan & Weinberg, 2011). La causa principal de ambos eventos genéticos y epigenéticos se debe al envejecimiento y la exposición a otros estímulos externos ambientales (Takeshima & Ushijima, 2019). Dentro de estas alteraciones, se pueden distinguir dos tipos eventos según su impacto en el fenotipo celular: conductores y pasajeros. Los eventos conductores, a diferencia de los segundos, son definidos como alteraciones que confieren una ventaja celular en términos de proliferación y supervivencia (McFarland et al., 2017) [Figura 3]. Estudios de secuenciación actuales concluyen que hasta el 97% de las mutaciones en el cáncer son eventos pasajeros (Lawrence et al., 2014). Aunque se piensa que estas alteraciones tan frecuentes poseen un efecto neutral o son deletéreas en la célula, en otros contextos podrían ralentizar el desarrollo tumoral o ser conductores latentes (McFarland et al., 2017; Nussinov & Tsai, 2015).



**Figura 3.** Selección de mutaciones genéticas y epigenéticas conductoras y pasajeras durante el desarrollo tumoral. Extraída de (Chatterjee et al., 2018).

Según el papel específico de los genes afectados por estos eventos genéticos en el contexto tumoral, se puede distinguir entre oncogenes y genes supresores tumorales. Los eventos que generan una ganancia de función en los oncogenes o causan la pérdida de función en los genes supresores tumorales pueden alterar el ciclo celular, otorgando así características asociadas al desarrollo de tumores. No obstante, cada vez existe más evidencia de que un mismo gen podría mostrar ambos comportamientos dependiendo del contexto celular (Zhu et al., 2015), como se explorará en los resultados del Capítulo IV.

## 1.2 Leucemia Linfoblástica Aguda

### 1.2.1 Definición, tipos e incidencia mundial

La leucemia linfoblástica se trata de una neoplasia hematológica cuyo origen se encuentra en la transformación maligna y en la proliferación de células progenitoras de la línea linfoide en la médula ósea, en la sangre y en otros lugares extramedulares del organismo (Terwilliger & Abdul-Hay, 2017). En función de la rapidez de proliferación de esta neoplasia, la leucemia linfoblástica puede ser clasificada como aguda (LLA) o crónica (LLC). A su vez, dependiendo del inmunofenotipo de la enfermedad, la transformación maligna puede afectar al linaje celular de los linfocitos B o de los linfocitos T, representando un 85% y un 15% de los casos totales aproximadamente de manera respectiva (Chennamadhavuni et al., 2023; Jędraszek et al., 2022). Gracias a los avances tecnológicos y a las nuevas metodologías de secuenciación, se han realizado estudios en grandes cohortes de pacientes con leucemia linfoblástica que han permitido detectar alteraciones genéticas para así conseguir una clasificación de hasta más de 30 subtipos moleculares con el fin de mejorar el diagnóstico del paciente y la estratificación para un tratamiento más adecuado (Brady et al., 2022).

La LLA es la neoplasia hematológica más común en el cáncer infantil, con un 80% de casos en pacientes pediátricos y un 20% en adultos aproximadamente (Terwilliger & Abdul-Hay, 2017). En 2019, la tasa de incidencia de la LLA fue 1,8 por cada 100.000 personas por año, mientras que la de mortalidad fue del 0,4 (Jędraszek et al., 2022). Siguiendo una distribución bimodal, la incidencia de LLA alcanza su mayor cifra en el período de edad comprendido entre los 1-4 años, luego disminuye drásticamente durante la infancia entre los 5 y 14 años, la adolescencia y adultos jóvenes, hasta alcanzar un segundo pico de incidencia en pacientes mayores de 45 años. (Jędraszek et al., 2022; PDQ Pediatric Treatment Editorial Board, 2002). Así, aproximadamente el 60% de los casos totales de LLA se diagnostican antes de los 20 años (Malard & Mohty, 2020).

## 1.2.2 Etiología

La LLA, al igual que otros tipos de cáncer, se origina debido a la interacción entre factores ambientales y una predisposición genética heredada. Uno de los principales desafíos en la investigación de esta enfermedad radica en reconocer las influencias ambientales significativas, las variaciones genéticas y comprender de qué manera y en qué momento estos elementos contribuyen al progreso de la leucemia.

### 1.2.2.1 Agentes ambientales externos

Aunque se han realizado numerosos estudios en leucemia infantil y adulta con el fin de correlacionar distintos factores ambientales y demográficos con el desarrollo de la enfermedad, las causas específicas siguen siendo aún desconocidas y las asociaciones frecuentemente son inconsistentes. Entre los factores externos más conocidos asociados a la etiología de la leucemia, se encuentra la exposición a radiación de iones o a sustancias químicas, por ejemplo, el benceno o pesticidas (Spector et al., 2006). Existen estudios que apuntan a la radiación ionizante como principal agente medioambiental externo que se correlaciona con el desarrollo de la leucemia, tanto en la preconcepción como en el período postnatal (Shu et al., 1994, 2002). En el caso de la exposición a campos electromagnéticos, existen publicaciones que indican que altas densidades de flujo están asociados a un pequeño aumento del riesgo con baja significancia (Inaba et al., 2013). Algunos agentes utilizados en quimioterapia, por ejemplo, los agentes alquilantes o los derivados de platino, suelen estar asociados con un mayor riesgo de padecer esta enfermedad (Tebbi, 2021). Aun dependiendo de la frecuencia y la cantidad, se ha sugerido que el consumo de tabaco y el alcohol de los padres, tanto en el período prenatal como postnatal de los niños, puede contribuir al desarrollo de la enfermedad (D. Pang et al., 2003). Por otro lado, existen estudios que respaldan la hipótesis de que exposiciones a infecciones en etapas tempranas y la adquisición adecuada de bacterias comensales del intestino pueden ayudar a disminuir el riesgo de padecer leucemia aguda (Greaves et al., 2021; Rudant et al., 2010).

Si bien la mayor parte de los agentes ambientales mencionados anteriormente muestran una ligera asociación con un mayor o menor efecto en el desarrollo de la leucemia aguda, aún se requieren estudios con modelos *in vitro* e *in vivo* que puedan confirmar los resultados observados.

### 1.2.2.2 Hipótesis sobre los factores causales del desarrollo de la LLA

En el siglo pasado, la principal causa etiológica propuesta para el origen de la leucemia infantil estaba enfocada en las infecciones. En 1988, se propusieron dos teorías principales basadas en este pretexto para explicar el desarrollo de esta enfermedad (Greaves, 2006). La primera de ellas, denominada mezcla poblacional (del inglés, “*population-mixing hypothesis*”), de Kinlen, establece que la leucemia es provocada por una respuesta anormal a una determinada infección común expandida debido a la migración de personas portadoras de la enfermedad entre distintas poblaciones. Por otro lado, la segunda hipótesis, denominada infección retardada (del inglés, “*delayed infection*”), de Greaves, sugiere que la leucemia infantil es causada por la ausencia o disminución de exposición a infecciones comunes durante la infancia. Como consecuencia, la falta de exposición resulta en un sistema inmunológico debilitado y en una respuesta deficiente ante infecciones bacterianas o virales. Aunque esta última hipótesis no se descarta la necesidad de pre-existencia de un posible clon preleucémico generado prenatalmente, la protección a través de la higiene que ofrece las sociedades más desarrolladas podría predisponer al sistema inmunológico a generar respuestas anómalas con el consiguiente desarrollo de la leucemia (Greaves, 2006; Law et al., 2003).

Aunque ambas hipótesis, la mezcla poblacional y la infección retardada, difieren en la base de los mecanismos patológicos, comparten que la leucemia infantil es una respuesta anormal del sistema inmunológico a infecciones comunes adquiridas en una situación sociodemográfica moderna. A nivel mecanístico, una respuesta inmunológica deficiente a través de células T podría provocar una respuesta inflamatoria exagerada y una liberación de citoquinas aberrante, que podrían afectar al correcto funcionamiento de las células madre hematopoyéticas (HSC, del inglés “*Hematopoietic Stem Cells*”) (Greaves, 2006). Si bien existen estudios que apoyan las hipótesis mencionadas anteriormente, aún no se ha llegado a un consenso definitivo sobre la relación etiológica de las infecciones y el desarrollo de la leucemia infantil.

### 1.2.2.3 Modelo del doble evento

Actualmente, se encuentra ampliamente aceptada la teoría del doble evento para el desarrollo de la LLA infantil, donde se postula que son necesarias dos etapas en las cuales ocurren lesiones genéticas. En la primera de ellas, se genera un clon pre-leucémico en el útero a través de la generación de diferentes alteraciones genéticas, por ejemplo, el gen de fusión *ETV6-RUNX1*. En segundo lugar, debido a diferentes factores externos ambientales, por ejemplo, una expo-



translocación t(12;21). En los últimos años, el diagnóstico molecular de la LLA ha sido complementado con técnicas transcriptómicas mediante el uso de *microarrays* para mejorar la precisión de la clasificación. Sin embargo, algunas de las alteraciones genéticas, tales como pequeñas deleciones o determinadas variaciones estructurales, siguen siendo complicadas de ser detectadas con el uso de *microarrays* (Coccaro et al., 2019). Actualmente, la implementación de técnicas de secuenciación masiva en la clínica, por ejemplo, el ARN-Seq, está permitiendo identificar nuevos biomarcadores de diagnóstico y posibles dianas terapéuticas en el marco de la medicina personalizada (Inaba et al., 2017).

### 1.2.4 Predisposición genética

La LLA se puede clasificar en diferentes subtipos genéticos, dependiendo del tipo de alteración presente en el genoma de las células afectadas. Estos diferentes tipos moleculares dependen principalmente de las alteraciones cromosómicas presentes, por ejemplo, las ganancias de copias de cromosomas completos o la reorganización de partes procedentes de éstos (Coccaro et al., 2019). Entre las posibles consecuencias producidas por estas alteraciones genéticas, se pueden encontrar el transporte de secuencias *enhancers* fuertes a otras partes del genoma o formar nuevos genes de fusión, que de manera muy frecuente implican genes de proteínas de la regulación de la transcripción, receptores de citoquinas o factores de transcripción hematopoyéticos. De manera secundaria, se pueden dar eventos relacionados con alteraciones en el número de copias de genes o mutaciones, sobre todo, en factores de transcripción específicos de los linfocitos. La identificación de estos subtipos en base a las alteraciones cromosómicas o génicas es relevante para la estratificación de los pacientes con LLA (Malard & Mohty, 2020; Roberts, 2018).

#### 1.2.4.1 Variaciones estructurales y numéricas de los cromosomas

Hasta ahora, los análisis citogenéticos realizados en muestras de pacientes con LLA han permitido detectar alteraciones estructurales y variaciones en el número de copias en el genoma relevantes para clasificar hasta el 80% de los afectados en subgrupos de pronóstico y tratamiento. Entre los biomarcadores más comunes, se pueden encontrar la alta hiperdiploidía (51-67 cromosomas), la hipodiploidía ( $\leq 44$  cromosomas), las fusiones génicas *BCR-ABL1* (cromosoma Philadelphia), *ETV6-RUNX1* y *TCF3-PBX1*, y las reorganizaciones del gen *KMT2A* (Malard & Mohty, 2020) [Figura 5]. Algunas de estas covariables clínicas descritas a continuación serán utilizadas en los análisis de supervivencia realizados en los Capítulos III y IV.

### 1.2.4.2 Variación en el número de cromosomas

La alta hiperdiploidía, que incluye la ganancia de al menos 5 cromosomas, se encuentra presente en el 25% de los pacientes con LLA, encontrándose asociada con un pronóstico favorable. (Malard & Mohty, 2020; Paulsson et al., 2015; Pui et al., 2019). La ganancia de cromosomas como aberración genómica presente en la LLA es considerada como uno de los eventos conductores primarios para el inicio de la neoplasia hematológica (Paulsson et al., 2015). Por el contrario, la hipodiploidía, caracterizada por presentar menos de 44 cromosomas, se encuentra asociada con un pronóstico desfavorable en los pacientes pediátricos con LLA. A pesar de desconocerse la ventaja selectiva que aporta la pérdida de los cromosomas, la presencia de este marcador en neoplasias hematológicas indica que está relacionada con la carcinogénesis. Dependiendo del número de cromosomas perdidos, esta anomalía se puede clasificar como hipodiploidía baja (32-29 cromosomas) o como casi haploidía (24-31 cromosomas). En el primero de los casos, hasta el 90% de los pacientes se caracterizan por poseer mutaciones con pérdida de función en el gen *TP53*, mientras que en la segunda categoría, existen alteraciones somáticas en genes relacionados con la señalización de los receptores de tirosina quinasa y RAS (Mullighan et al., 2015; Safavi & Paulsson, 2017).

### 1.2.4.3 Translocaciones cromosómicas

Las translocaciones cromosómicas son utilizadas de manera frecuente para estratificar los pacientes con el fin de derivarlos al tratamiento más adecuado, debido a su asociación significativa con el pronóstico. Estas alteraciones genéticas suelen implicar la interrupción de genes relacionados con la regulación de la hematopoyesis, la activación de oncogenes o la inhibición de genes supresores tumorales (Malard & Mohty, 2020).

- **Translocación t(12;21)(p13;q22) – *ETV6-RUNX1* (*TEL-AML1*)**

La reorganización cromosómica más común en LLA-B pediátrica es la translocación t(12;21)(p13;q22), que genera el gen de fusión *ETV6-RUNX1* (Inaba & Mullighan, 2020). La fusión provocada por esta translocación está compuesta por dos genes reguladores críticos de la hematopoyesis: la porción 5' del gen *ETV6* (*TEL*) del cromosoma 12p13 y prácticamente el gen *RUNX1* (*AML1*) completo del cromosoma 21q22. El gen *ETV6* se encuentra altamente expresado en HSC como factor de transcripción de la familia ETS (del inglés, "*Erythroblast Transformation Specific*"), implicado en la regulación de la proliferación y la diferenciación de estas células progenitoras, la angiogénesis y el desarrollo de megacariocitos (Bohlander, 2005). De igual manera, *RUNX1* es otro factor de

transcripción involucrado en la regulación del proceso de la hematopoyesis. (Grossmann et al., 2011). El transcrito de la fusión *ETV6-RUNX1* es detectado aproximadamente hasta en el 25% de los pacientes pediátricos con LLA de subtipo pre-B (Kaczmarska et al., 2023). Gracias a análisis realizados en sangre de cordón umbilical, se ha propuesto que la fusión oncogénica *ETV6-RUNX1* es un evento inicial prenatal no suficiente para iniciar el desarrollo tumoral (Schäfer et al., 2018). Así pues, se ha propuesto esta translocación genera un primer clon preleucémico que necesitará eventos secundarios genéticos postnatales para iniciar la leucemia (Isidro-Hernández et al., 2022).

Según los resultados obtenidos hasta la fecha, la t(12;21) se considera un indicador pronóstico favorable asociado con mejores resultados en el tratamiento de los pacientes. Gracias a los estudios de secuenciación actuales, se han descubierto subtipos similares a *ETV6-RUNX1* en cuanto a su perfil de expresión génica y los inmunofenotipos. Entre los genes afectados por una variación en su número de copias o por translocaciones, se encuentran los pertenecientes a la familia de factores de transcripción ETS, IKZF1 o TCF3. De manera similar a la translocación t(12;21), estos subtipos se encuentran relacionados con un pronóstico relativamente favorable de los pacientes (Inaba & Mullighan, 2020).

- **Translocación t(1;19)(q23;p13) – TCF3-PBX1 (E2A-PBX1)**

La translocación t(1;19)(q23;p13) genera el gen de fusión *TCF3-PBX1*, considerado como un factor de transcripción oncogénico, asociada con el inmunofenotipo del subtipo pre-B (Harrison et al., 2017). Se trata de una reorganización cromosómica recurrente que se encuentra presente como subtipo principal en aproximadamente hasta el 7% de los pacientes pediátricos con LLA (Isidro-Hernández et al., 2022). El gen de fusión generado codifica un factor de transcripción que contiene el dominio de transactivación de TCF3 junto al dominio de unión a ADN de PBX1, que representa el evento inicial de esta enfermedad (Tirado et al., 2015). Se ha demostrado que la activación condicional de este factor de transcripción anormal en células B progenitoras produce un aumento de la auto-renovación de estas células junto a otras alteraciones genéticas, por ejemplo, la delección de *PAX5* o la activación de la señalización de la ruta JAK/STAT (Dunque-Afonso et al., 2015).

Tradicionalmente, el subtipo de LLA con t(1;19) ha sido clasificado como un factor de riesgo alto para los pacientes debido a la alta afectación del sistema nervioso central (SNC). Sin embargo, gracias a tratamientos intensivos de quimioterapia con altas dosis de metotrexato, la presencia de este gen de fusión está asociada con un pronóstico favorable o de riesgo intermedio (Harrison et al., 2017). No obstante, en otras posibles variaciones menos frecuentes de esta trans-

locación, por ejemplo, en la *TCF3-HLF*, el pronóstico de los pacientes sigue siendo extremadamente desfavorable (Gu et al., 2019; Harrison et al., 2017).

- **Translocación t(9;22)(q34;q11.2) – *BCR-ABL1***

El cromosoma Philadelphia (Ph) se define por la presencia de la fusión génica *BCR-ABL1*, que se encuentra presente en aproximadamente hasta el 5% de los pacientes pediátricos y el 25% de los pacientes adultos con LLA-B (Mullighan, 2014). *BCR-ABL1* se define como una quinasa oncogénica con actividad constitutiva, cuyas isoformas principales específicas de enfermedad, p190 y p210, se definen dependiendo del punto de ruptura en el gen *BCR*. Mientras que la isoforma p210 se encuentra asociada a la leucemia mieloide crónica (LMC), la isoforma p190 es característica de LLA (Adnan-Awad et al., 2021). La translocación t(9;22)(q34;q11.2) se encuentra relacionada con alteraciones genéticas que provocan una desregulación de la señalización de los receptores de citoquina y las tirosin-quinasas. Entre los genes afectados, se puede encontrar reorganizaciones de *CRLF2* y *ABL*, mutaciones inactivadoras de la ruta de JAK-STAT o alteraciones en otras quinastas, por ejemplo, *NTRK3*. Los anteriores genes mencionados se encuentran involucrados en la regulación del ciclo celular, la diferenciación y la proliferación de las células B (Hunger & Mullighan, 2015b; Roberts, 2017).

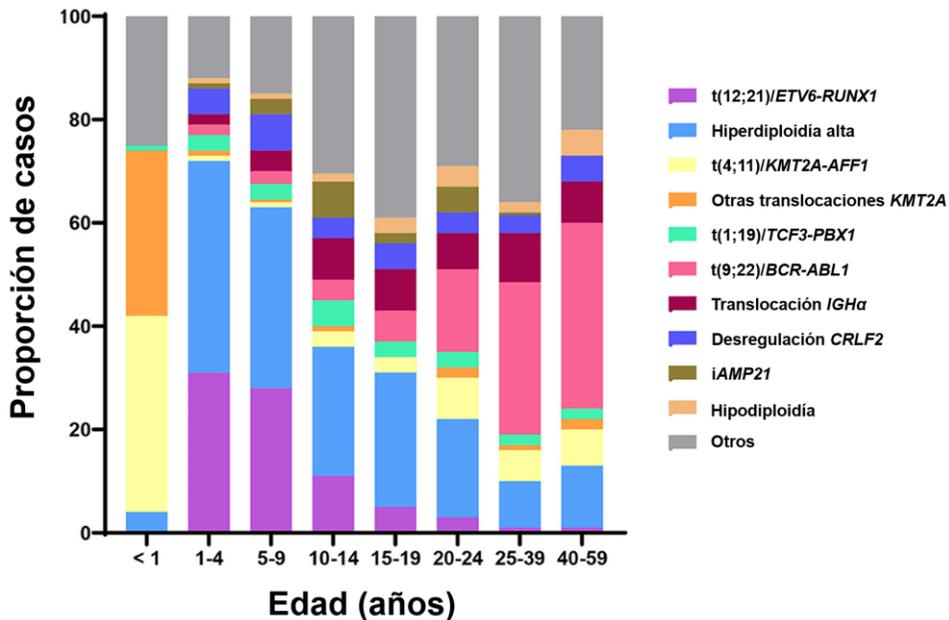
Los pacientes positivos para el cromosoma Ph poseen un incremento de riesgo de afectación del SNC, un curso clínico agresivo de la enfermedad y una alta tasa de recidiva (Liu-Dumlao et al., 2012). Dado que la gran mayoría de quinastas identificadas en presencia del cromosoma Ph convergen generalmente en las vías de señalización mencionadas anteriormente, el tratamiento de esta enfermedad puede verse mejorado con la combinación de la quimioterapia intensiva con inhibidores de tirosin-quinastas, por ejemplo, de manera específica para *ABL1* o *JAK* (Roberts, 2017). En el futuro, los ensayos clínicos y el estudio de los mecanismos de resistencia a estos tratamientos mejorarán el pronóstico desfavorable del cromosoma Ph en pacientes pediátricos y adultos de LLA-B.

- **Reordenamientos del gen *KMT2A* (*MLL*)**

El gen *KMT2A*, que codifica a la metiltransferasa específica de lisina [K]-2A, es conocido tradicionalmente como el gen de la leucemia del linaje mixto (*MLL*, del inglés “*Mixed-Lineage Leukemia*”). La función biológica principal de este gen es la regulación epigenética del inicio y elongación de la transcripción a través de la metilación de H3K4 de sus genes diana (Górecki et al., 2023). El reordenamiento del gen *KMT2A* se encuentra presente en aproximadamente el 5% de los pacientes pediátricos y el 10% de adultos con LLA, presentando tanto

características fenotípicas linfoides como mieloides (El Chaer et al., 2020). A nivel celular, se encuentra implicado en la regulación de la proliferación y la diferenciación durante la hematopoyesis (Steinhilber & Marschalek, 2018). En relación con el desarrollo de neoplasias, el gen *KMT2A* es uno de los puntos más probables de recombinación en el genoma humano, conociéndose una gran variedad de alteraciones genómicas, tales como translocaciones o duplicaciones en tándem internas. Aunque la frecuencia de las distintas fusiones génicas posibles de este gen puede variar dependiendo de la edad del paciente, se puede encontrar entre las más recurrentes las producidas con los genes *AFF1*, *MLLT1* y *MLLT3* (Górecki et al., 2023).

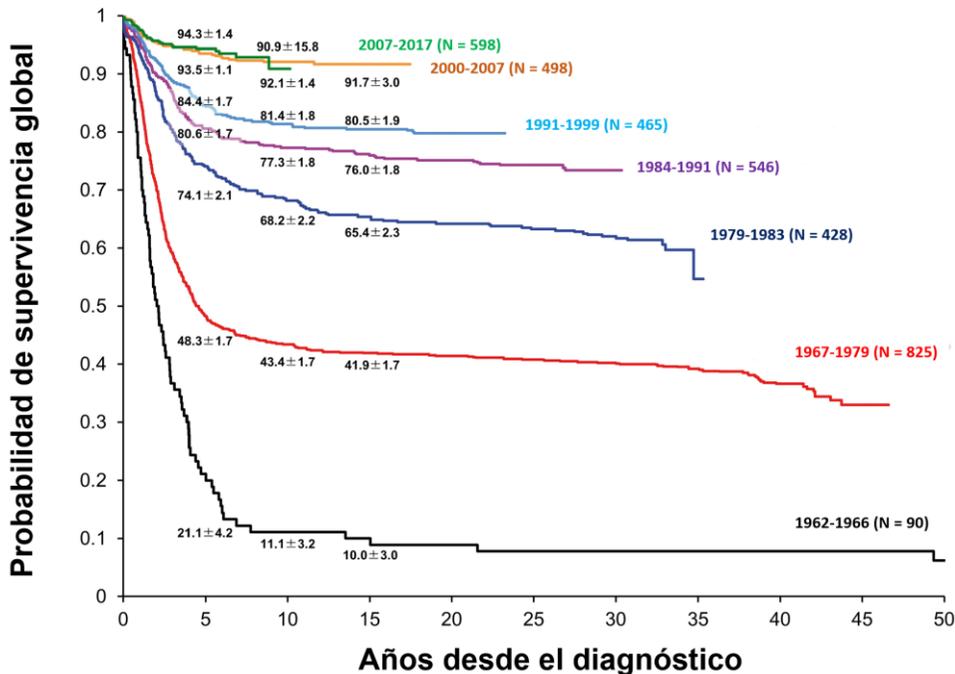
La presencia del reordenamiento del gen *KMT2A* generalmente es considerada un factor pronóstico desfavorable independiente, caracterizado por una tendencia alta a la recaída (Górecki et al., 2023; Tauchi et al., 2008). Dado que los pacientes con LLA que presentan este reordenamiento no son sensibles a las terapias estándar, se han diseñado fármacos que interrumpen la interacción de la metiltransferasa *KMT2A* con otras proteínas. Por ejemplo, el uso de peptidomiméticos evitan la interacción de *KMT2A* y *WDR5* inhibiendo la expresión de genes aguas abajo que podrían ser responsables del inicio de la leucemia (Karatat et al., 2013).



**Figura 5.** Distribución de subtipos citogenéticos de LLA-B en diferentes rangos de edad. Adaptación de (Marcotte et al., 2021).

## 1.2.5 Tratamientos

A lo largo de los últimos 50 años, la mejora de los protocolos quimioterapéuticos ha permitido alcanzar una remisión de la enfermedad en un porcentaje considerablemente alto de los pacientes pediátricos [Figura 6]. Dentro de estos avances, se puede encontrar la adición de la administración intratecal, la radioterapia craneal o el control de las dosis de los agentes quimioterapéuticos (Hunger & Mullighan, 2015a). Actualmente, el tratamiento de la LLA consta de tres etapas principales: inducción de la remisión (4-6 semanas), consolidación o intensificación (6-8 meses) y mantenimiento (18-30 meses). En la inducción, se usan medicamentos como glucocorticoides, vincristina y asparraginasas, a veces con antraciclinas para prevenir infecciones. En la segunda etapa de consolidación, se utiliza quimioterapia intensiva y ácido fólinico para proteger los tejidos normales. Por último, en la fase de mantenimiento, se administran agentes antimetabólicos de baja intensidad para retrasar el crecimiento de las células tumorales, entre los que se puede encontrar el metotrexato o la mercaptopurina. (Chang et al., 2021; Hunger & Mullighan, 2015a; Inaba & Mullighan, 2020).



**Figura 6.** Mejora de la supervivencia global de los pacientes con LLA pediátricos tratados en distintas cohortes de estudio del St. Jude Children's Research Hospital. Adaptación de (Inaba & Mullighan, 2020).

El porcentaje de recaída de los pacientes pediátricos con LLA puede alcanzar hasta el 20% de los casos, siendo la posibilidad de la remisión completa aún más complicada tras este evento (Yu et al., 2020). Las células leucémicas derivados de estos casos suelen poseer mutaciones que disminuyen la sensibilidad al tratamiento quimioterapéutico estándar establecido (X. Ma et al., 2015; Mullighan et al., 2008). El descubrimiento de la aplicación de nuevas terapias dirigidas e inmunoterapias combinadas ha mejorado considerablemente el porcentaje de pacientes pediátricos curados, por ejemplo, los inhibidores de tirosina quinasa (Malczewska et al., 2022). En el caso de aquellos pacientes estratificados como alto riesgo o aquellos que poseen la enfermedad refractaria, el alotrasplante de progenitores hematopoyéticos ha sido considerado el procedimiento estándar para conseguir una remisión duradera (Terwilliger & Abdul-Hay, 2017). Entre los factores para considerar esta opción en los pacientes de alto riesgo, se encuentran la presencia del cromosoma Ph, la afectación del SNC, el recuento elevado de glóbulos blancos, la hipodiploidía y otras translocaciones con pronóstico desfavorable (Dombret et al., 2002; Jamieson et al., 2003).

## **1.3 Regulación de la expresión génica**

### **1.3.1 Epigenética**

El término “epigenética” se define clásicamente como aquellos cambios heredables en la expresión génica que no se deben directamente a alteraciones en la secuencia de ADN. Los cambios epigenéticos son esenciales para la diferenciación y el crecimiento de los diferentes tipos celulares dentro de un organismo (Biswas & Rao, 2018; Eccleston et al., 2007). En su conjunto, las modificaciones epigenéticas presentes en una célula definen el epigenoma celular, cuyo estado se encuentra en continuo cambio en respuesta y adaptación a las señales del medio ambiente (Cavalli & Heard, 2019). Esta información epigenética celular puede estar representada de distintas formas: desde modificaciones directas en el ADN o en las histonas, así como una regulación ejercida por proteínas o ARN no codificantes (Eccleston et al., 2007).

Sin embargo, una alteración de las marcas epigenéticas puede conducir a una activación o una inhibición inapropiada de la expresión génica desembocando en cambios patológicos, por ejemplo, en el desarrollo tumoral (Kurdistani, 2011; H. Shen & Laird, 2013). Así, las modificaciones epigenéticas consideradas como principales mecanismos afectados durante la progresión del cáncer se pueden clasificar en tres grupos: metilación del ADN, modificaciones de histonas y ARN no codificantes (Lu et al., 2020).

## 1.3.2 Mecanismos de regulación de la expresión génica

### 1.3.2.1 Metilación del ADN y el ARN

La metilación del ADN es uno de los principales mecanismos epigenéticos que regula funcionalmente una gran variedad de procesos biológicos, tales como la inactivación del cromosoma X, la impronta genómica o la programación de la identidad celular (Gopalakrishnan et al., 2008; Robertson, 2005). La metilación se produce en la posición 5 de los anillos de pirimidina de las citosinas, principalmente en unas regiones localizadas en la región 5' del promotor de más del 50% de los genes humanos denominadas islas CpG (Lister et al., 2009). En general, la adición covalente de grupos metilo impide la unión de factores de transcripción o atrae proteína específicas para reconfigurar la cromatina, provocando por lo general en consecuencia una represión de la expresión génica (Lu et al., 2020).

El control de la metilación del ADN es esencial para el desarrollo adecuado del organismo (Jaenisch & Bird, 2003). En el contexto tumoral, se ha observado una desregulación de este mecanismo epigenético, donde la hipometilación de regiones metiladas de ADN puede provocar la activación transcripcional de oncogenes, así como una inestabilidad genómica. En contraposición, la hipermetilación ha sido tradicionalmente asociada al silenciamiento de una gran variedad de genes que controlan la proliferación celular, la apoptosis o la diferenciación, considerados como supresores tumorales. No obstante, existen situaciones más complejas donde la hipometilación puede conducir a represión de genes, y viceversa (Berman et al., 2012; Laird-Offringa & Sanchez-Cespedes, 2018).

Respecto a las modificaciones que puede recibir el ARN, la metilación de la posición N-6 del residuo adenosina ( $m^6A$ ) se encuentra presente principalmente dentro de exones largos, en la región 3' UTR y cerca de los codones de STOP (Ke et al., 2015; Meyer et al., 2012). La  $m^6A$  se trata de una modificación reversible y dinámica que puede influir en la transcripción, la traducción, la degradación o el *splicing* del ARN (X.-Y. Chen et al., 2019).

### 1.3.2.2 Modificaciones de histonas

La unidad principal de la cromatina, conocida como nucleosoma, se encuentra compuesto por un octámero de proteínas histonas que envuelve el ADN. Esta estructura está formada por un tetrámero de un par de histonas 2A (H2A) y un par de histonas 2B (H2B), rodeado por dímeros de histona 3 (H3) e histona 4

(H4). Las regiones N-terminal ricas en lisina y en arginina de las histonas se encuentran extendidas más allá del núcleo, encontrándose sujetas a modificaciones epigenéticas, tales como ubiquitinación, acetilación, fosforilación, metilación, sumoilación, citrulinación y otras existentes. Estos cambios epigenéticos pueden alterar las interacciones electroestáticas entre los residuos cargados positivamente de las histonas y los grupos fosfatos del esqueleto del ADN, regulando así el acceso de otras proteínas a la cromatina (Lu et al., 2020). Las modificaciones son realizadas por una red interactiva de ADN metiltransferasas, enzimas modificadoras de histonas y otros factores interactores de la cromatina (Adcock et al., 2006).

Entre las modificaciones de las histonas más estudiadas que juegan un papel importante en la regulación transcripcional, se encuentran la acetilación y la metilación de residuos de lisina de H3 o H4. En general, la adición de grupos acetilo puede descompactar la cromatina permitiendo la entrada de factores de transcripción para activar la expresión génica (Lu et al., 2020). En contraposición, el efecto producido por la metilación depende de los residuos afectados de las histonas. Así, la metilación de la lisina 4 ó 36 de H3 puede conducir a una activación transcripcional, mientras que la metilación de la lisina 20 de la H4 es reconocida por ser una marca epigenética represiva (Barski et al., 2007; Wiles & Selker, 2017). Debido a la compleja red de regulación epigenética, las alteraciones de las enzimas que regulan estas marcas pueden conducir a modificaciones epigenéticas anómalas presentes en enfermedades (Laird-Offringa & Sanchez-Cespedes, 2018).

### 1.3.2.3 Compactación de histonas

La estructura y la accesibilidad de la cromatina no solo se encuentra reguladas por enzimas modificadores de histonas, sino también por complejos remodeladores de la cromatina dependientes de ATP que actúan como moduladores de la arquitectura genómica (Singh et al., 2023). Estos complejos multiproteicos poseen la capacidad de regular el acceso al ADN al reorganizar, expulsar, deslizar o modificar los nucleosomas. Así, una vez han alterado el contacto entre histona y ADN, proporcionan acceso a otras proteínas que pueden interactuar con ambas estructuras (Magaña-Acosta & Valadez-Graham, 2020).

Existen cuatro familias de complejos remodeladores de la cromatina distinguibles por la subunidad ATPasa que la componen: SWI/SNF, INO80, ISWI y CHDs (Bartholomew, 2014). Aunque cada una de ellas interactúan con los nucleosomas de una manera específica, comparten algunas características comunes: (I) poseen mayor afinidad por el nucleosoma que por el ADN liberado; (II) contiene una subunidad catalítica única que contiene un dominio ATPasa; (III) y

tienen dominios o proteínas que pueden regular la actividad ATPasa y capaces de interactuar con otras proteínas presentes en la cromatina, tales como factores de transcripción. La tercera característica mencionada es la que permite a estas familias de complejos remodeladores de la cromatina actuar de manera específica en determinados nucleosomas de ubicaciones genómicas particulares (Clapier et al., 2017). En el apartado 1.4, se detallará la familia de complejos remodeladores de la cromatina SWI/SNF estudiada en esta tesis doctoral.

### 1.3.2.3 ARN no codificantes

Los ARN no codificantes son moléculas no traducidas a proteínas que desempeñan un papel importante en el control de la expresión génica y proteica. Dependiendo de sus funciones dentro de las células, pueden intervenir en tres contextos principales: regulación de la expresión génica (por ejemplo, los micro ARN (miARN) y los ARNlnc), síntesis de proteínas (por ejemplo, los ARN ribosómicos (ARNr)) y síntesis de ARN (por ejemplo, los ARN pequeños nucleares). En el contexto de la regulación epigenética, la desregulación de estos ARN no codificantes puede desencadenar el desarrollo de enfermedades. Entre los mecanismos de regulación más relevantes, destaca el proceso de silenciamiento post-transcripcional llevado a cabo por secuencias cortas de ARN de aproximadamente 22 pares de bases conocidos como miARN. Estas moléculas se unen específicamente a sus dianas mediante complementariedad de bases, impidiendo así la traducción de los ARN mensajeros (ARNm). Además, los miARN pueden ser objeto de modificaciones químicas que podrían afectar a su función en la regulación de la expresión génica (Pathania et al., 2022; Yao et al., 2019).

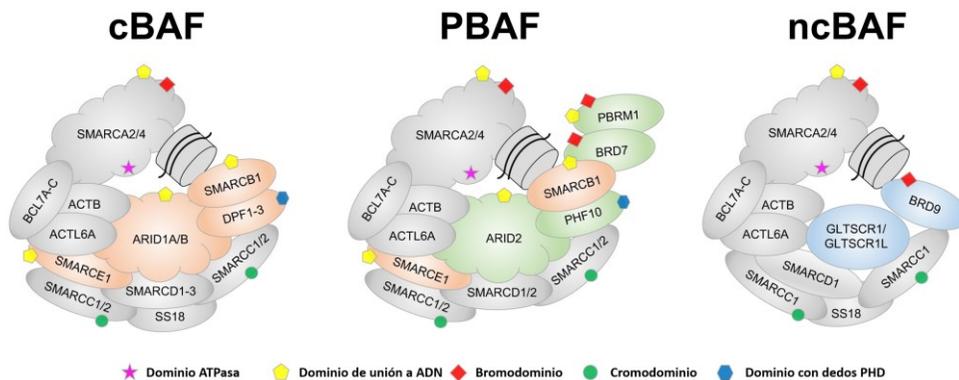
Otras de las moléculas de más reciente descubrimiento son los ARNlnc, cuya definición y funciones en la regulación de la expresión génica se explorarán en mayor profundidad en el apartado 1.5.

## 1.4 Complejo regulador de la cromatina SWI/SNF

En la última década, el complejo regulador de la cromatina SWI/SNF, que fue previamente descubierto en *Saccharomyces cerevisiae* y *Drosophila*, ha sido estudiado estructural y funcionalmente en profundidad en mamíferos (Singh et al., 2023). Este complejo multiprotético emplea la energía derivada de la hidrólisis del ATP para desplazar a los nucleosomas a lo largo del ADN, expulsarlos o alterar su composición. En consecuencia, pueden regular el acceso a la cromatina de una gran cantidad de moléculas, tales como maquinaria y factores

de transcripción, proteínas de unión a ADN, u otros cofactores, con el fin de controlar la expresión génica (Z. Li et al., 2023).

Actualmente, se han identificado 29 genes que codifican para distintas subunidades SWI/SNF. Sin embargo, solo se han identificado aproximadamente hasta 12 subunidades coexistiendo en un mismo complejo funcional. A su vez, los complejos SWI/SNF en mamíferos se pueden dividir en tres subfamilias: complejo cBAF (del inglés, “*Canonical BRG1-associated Factor*”), complejo PBAF (del inglés, “*Polybromo BRG1-Associated Factor*”) y el complejo ncBAF (del inglés, “*Non-canonical BRG1-Associated Factor*”). Las tres subfamilias mencionadas contienen un grupo de subunidades de ensamblaje que conforman el núcleo del complejo: SMARCC1, SMARCC2, SMARCD1, SMARCD2, SMARCD2, y alguna de las dos subunidades ATPasas mutuamente exclusivas, SMARCA4 o SMARCA2. Además, existe un grupo de subunidades accesorias que pueden caracterizar las tres subfamilias previamente mencionadas, por ejemplo, las subunidades ARID1A/ARID1B en el complejo cBAF o la subunidad ARID2 y la subunidad PBRM1 en el complejo PBAF (Andrades et al., 2023; Z. Li et al., 2023). **[Figura 7]** La composición del complejo SWI/SNF determina de manera específica el lugar de acción en el genoma y su interacción con distintos factores de transcripción (Andrades et al., 2023).



**Figura 7.** Estructura del ensamblaje de las tres subfamilias de complejos remodeladores de la cromatina SWI/SNF. En color gris, se representa las subunidades compartidas entre los tres subcomplejos; en color naranja, aquellas exclusivas de los BAF que no se encuentran en la subfamilia ncBAF; en color verde, aquellas específicas del PBAF; en color azul, aquellas específicas del ncBAF. Adaptada de (Wanior et al., 2021).

El papel del complejo remodelador de la cromatina en el desarrollo tumoral ha sido analizado en numerosos estudios de secuenciación. Los genes codificantes de las subunidades del complejo SWI/SNF se encuentran alterados hasta en el 25% de los cánceres humanos (Wanior et al., 2021). La alta tasa mutacional descrita en las subunidades del complejo SWI/SNF indica un posible rol supresor

tumoral, aunque los efectos funcionales que derivan de estas alteraciones genéticas en distintas subunidades para la misma enfermedad pueden ser diferentes (Andrades et al., 2023). Por ello, se plantea la cuestión de cómo las alteraciones genéticas en los complejos remodeladores de la cromatina SWI/SNF contribuyen a la formación de tumores, cómo impulsan el crecimiento de los mismos y, lo que es más importante, si es posible utilizarlos como diana terapéutico (Wanior et al., 2021). Específicamente, en esta tesis doctoral se presentan resultados relacionados con las alteraciones de las subunidades ARID1A y ARID1B, que forman parte del complejo cBAF, en el capítulo V; y de ARID2, que forma parte del complejo PBAF, en el capítulo IV.

## 1.4.1. cBAF: Las subunidades ARID1

### 1.4.1.1 Definición y estructura de ARID1A y ARID1B

ARID1A/BAF250A (del inglés, “*AT-rich interaction domain 1A*”) y ARID1B/BAF250B (del inglés, “*AT-rich Interaction Domain 1B*”) son dos subunidades mutuamente exclusivas de la subfamilia de complejos remodeladores de la cromatina SWI/SNF denominada cBAF. La longitud de la secuencia proteica de ARID1A es de 2.285 aminoácidos, mientras que la de ARID1B es un poco menor con 2.236 residuos. Si bien los primeros 600 residuos en el extremo N-terminal de ambas secuencias son muy divergentes, comparten aproximadamente un 50% de identidad en su composición. Además de compartir el dominio ARID de unión a ADN con un 80% de homología, ambos poseen el dominio BAF250\_C para poder interactuar con otras proteínas del complejo (Van Rechem et al., 2009; Wu & Roberts, 2013). Por otro lado, se puede encontrar en su estructura otros motivos relevantes, tales como el LXXLL, el de caja BC y señales de localización nuclear (Bateman et al., 2016; X. S. Li et al., 2010). De manera diferencial, ARID1A contiene un dominio adicional de unión a HIC1 implicado en la regulación de la transcripción (Odnokoz et al., 2021).

Una vez formado el núcleo en el ensamblaje del complejo remodelador de la cromatina cBAF, ARID1A y ARID1B se pueden adherir para unir las subunidades accesorias DPF (DPF1, DPF2, DPF3) e incorporar el módulo ATPasa. Para ello, según los estudios realizados en este proceso de ensamblaje, ambas subunidades poseen regiones específicas en el extremo C-terminal denominadas CBR (del inglés, “*Core Binding Region*”) y R4. La primera de ellas muestra mayor afinidad por unirse a las subunidades SMARCC y SMARCD, mientras que la segunda podría interactuar con la ATPasa SMARCA y la subunidad ACTL6A. A diferencia de estas regiones del extremo C-terminal, el dominio ARID de am-

bas subunidades se encuentra implicado mayormente en el reclutamiento del complejo cBAF en el ADN (Mashtalir et al., 2018).

#### 1.4.1.2 Roles fisiológicos celulares de las proteínas ARID1

Dado que las proteínas ARID1 forman parte de la estructura del complejo remodelador de la cromatina SWI/SNF, deben jugar un papel importante en numerosos procesos esenciales celulares, tales como la regulación transcripcional, la reparación del ADN o la proliferación celular (Masliah-Planchon et al., 2015; Pulice & Kadoch, 2016). Aun siendo ARID1A y ARID1B similares a nivel estructural en sus dominios, las funciones que ambas realizan de manera específica aún se encuentran por dilucidar. Así pues, se requiere explorar en profundidad los factores celulares que regulan la incorporación de ARID1A o su parálogo ARID1B al complejo remodelador de la cromatina cBAF para ejercer estas funciones dependientes de contexto (Odnokoz et al., 2021).

Las proteínas ARID regulan la accesibilidad de la cromatina a través de los complejos remodeladores a través de su unión a ADN de manera no dependiente. De esta manera, pueden regular el control del acceso de la ARN polimerasa II en los sitios de inicio de transcripción (TSS, del inglés “*Transcription Start Site*”) con el fin de controlar la expresión génica (Trizzino et al., 2018). Ambas subunidades pueden interactuar con reguladores transcripcionales específicos o ser reclutadas en promotores para colaborar en la activación o represión de genes (Odnokoz et al., 2021). Por ejemplo, el factor de transcripción HIC1 puede interactuar con ARID1A para ser reclutado en promotores específicos con el fin de reprimir la transcripción de algunos genes, tales como *ATOH1* o *E2F1* (Van Rechem et al., 2009). ARID1A también puede inhibir directamente la interacción de la ARN polimerasa II con el TSS del promotor de *TERT*. En este último caso, si ARID1A se encuentre alterada genéticamente y no es funcional, la reactivación transcripcional de *TERT* puede provocar una mayor supervivencia de las células a través del mantenimiento de los telómeros en un contexto tumoral (Rahmanto et al., 2016). Por otro lado, las proteínas ARID1 también pueden interactuar con enzimas modificadoras de histonas para regular la expresión génica, tales como histonas deacetilasas (HDAC, del inglés “*Histone Deacetylase*”) o histonas acetiltransferasas (HAT, del inglés “*Histone Acetyltransferase*”) (Odnokoz et al., 2021).

En el contexto de la reparación de ADN dañado, se ha visto que las proteínas ARID1 pueden atraer los complejos remodeladores de la cromatina cBAF para realizar la reparación de ADN por recombinación homóloga y mediante la unión de extremos no homólogos en la rotura de doble cadena. Así, se ha demostrado que ARID1A desempeña un papel importante interactuando con distintas

proteínas para completar este proceso (X. Wang et al., 2014). Por ejemplo, ARID1A puede interactuar con las quinasas de punto de control de daño en el ADN ATM y ATR con el fin de regular la reparación de roturas de doble cadena (Jiang et al., 2019; J. Shen et al., 2015). Adicionalmente, ARID1A interviene en la mitosis para estabilizar cromátidas hermanas a través de su interacción con TOP2a (Dykhuisen et al., 2013).

### 1.4.1.3 Alteraciones de las proteínas ARID1 en cáncer

Las alteraciones genéticas en ARID1A y ARID1B pueden afectar a la regulación transcripcional de diferentes maneras dependiendo del contexto celular y de cómo se encuentren afectados los genes. En todos los tipos de cáncer, ARID1A y ARID1B contienen mutaciones somáticas hasta aproximadamente en el 8% y 3% de las muestras, respectivamente. Las mutaciones suelen ubicarse en el extremo C-terminal de las proteínas en los dominios ARID y BAF250\_C. En consecuencia, la interacción con el resto de las subunidades del complejo remodelador de la cromatina cBAF y la interacción con el ADN para regular la expresión génica pueden verse alteradas (Mashtalir et al., 2018; Odnokoz et al., 2021; Sandhya et al., 2018). Las mutaciones para ambos genes frecuentemente son sin sentido o cambio en el marco de lectura (Pagliaroli & Trizzino, 2021). En cuanto a los niveles de expresión génica, ARID1A se encuentra mayormente expresado que ARID1B en una gran variedad de tejidos. Por ello, se plantea si la pérdida de ARID1A puede ser compensada por la de su parólogo, o si de manera no excluyente ARID1B también ejerce funciones específicas celulares (Pulice & Kadoch, 2016; X. Wang et al., 2004).

ARID1A es la subunidad más frecuentemente mutada de los miembros del complejo SWI/SNF en cáncer, incluyendo cáncer de ovario, cáncer pancreático, carcinoma renal, cáncer de mama y cáncer de pulmón. Actualmente, se plantea el posible papel dual de ARID1A como supresor tumoral y gen oncogénico (Lebedev et al., 2023). En las muestras tumorales donde ambas subunidades se encuentren alteradas y la expresión proteica sea deficiente, se produce la completa desestabilización del complejo SWI/SNF y una proliferación celular deteriorada (Helming et al., 2014). En cambio, las alteraciones genéticas de ARID1B se encuentran frecuentemente asociadas a trastornos del neurodesarrollo, tales como el síndrome de Coffin-Siris, trastornos del espectro autista o neuroblastoma (Smith et al., 2020; van der Sluijs et al., 2019). Dada la diferente tasa mutacional y la presencia de estas alteraciones en diferentes enfermedades, se sugiere que posiblemente ambas subunidades posean funciones específicas en diferentes líneas celulares.

## 1.4.2 PBAF: La subunidad ARID2

### 1.4.2.1 Definición y estructura de ARID2

ARID2 (del inglés, “*AT-rich interaction domain 2*”), o también conocida como BAF200, es una proteína componente específica del complejo remodelador de la cromatina PBAF, cuya secuencia proteica se encuentra compuesta por 1.835 aminoácidos, codificada por el gen *ARID2* en el cromosoma 12q (Loesch et al., 2020). ARID2 fue descubierto como la proteína zipzap/p200 que interactúa con factores de respuesta al suero para controlar transcripcionalmente la expresión génica cardíaca (X. Zhang et al., 2006). A nivel estructural, ARID2 se caracteriza por poseer un dominio de interacción ARID en su extremo N-terminal, y otros dos adicionales de unión a ADN: un dominio RFX (del inglés, “*Regulatory Factor X*”) conservado a nivel evolutivo y un dominio de dedos de zinc C2H2 en el extremo C-terminal. Adicionalmente, existe un dominio rico en glutamina y prolina, y un dominio LXXLL, en el medio de la secuencia aminoacídica de ARID2, relacionados con una posible función de interacción proteína-proteína (Loesch et al., 2020).

Según los estudios realizados sobre el proceso de ensamblaje del complejo remodelador de la cromatina PBAF, la construcción de este complejo es iniciada por ARID2 constituyendo un puente sólido entre el módulo ATPasa y la estructura nuclear. La delección de esta proteína supone una pérdida de estabilidad para la incorporación de otras subunidades importantes del complejo PBAF, tales como BRD7, PBRM1 y PFH10. Por ello, se presupone que ARID2 es importante para el reclutamiento y la estabilidad del resto de subunidades que componen este sistema (Mashtalir et al., 2018).

### 1.4.2.2 Funciones fisiológicas de ARID2

ARID2, como uno de los componentes principales del complejo remodelador de la cromatina PBAF, se encuentra estrechamente relacionado con la regulación de la expresión génica tejido específica (X. Zhang et al., 2006). Existen estudios donde se ha descrito la participación de ARID2 en el control del ciclo y la proliferación celular a través de su interacción con el factor de transcripción E2F1 para inactivando las ciclinas D1 y E1 (Duan et al., 2016). En el contexto de las células madre, ARID2 está implicado en la regulación de la embriogénesis y el desarrollo de órganos, por ejemplo, en la morfogénesis del corazón y en la angiogénesis de la arteria coronaria (He et al., 2014), o en la diferenciación de los osteoblastos (Xu et al., 2012). Adicionalmente, se encuentra involucrado en la regulación de la respuesta al daño en el ADN, ya que su delección provoca la desregulación de la reparación por escisión de nucleótidos en células humanas de

hepatocarcinoma (Oba et al., 2017). De acuerdo con esta última función de ARID2, también se ha descubierto que facilita los procesos de recombinación homóloga, por ejemplo, interaccionando con RAD51 a través del dominio carboxiterminal para la reparación de rotura de doble cadena en el ADN (de Castro et al., 2017). Por último, el complejo PBAF, caracterizado por la especificidad que proporciona la subunidad ARID2, se encuentra relacionado con la activación de genes esenciales de la primera fase de la meiosis en machos. A nivel molecular, ARID2 puede estar implicado en el ensamblaje del huso mitótico, la maduración del centrosoma y la regulación de la cantidad de fosfatasa con el fin de que la meiosis progrese adecuadamente (Menon et al., 2021).

### **1.4.2.3 ARID2 en el contexto tumoral.**

Las alteraciones genéticas de ARID2 han sido detectadas en distintos tipos de cáncer, principalmente en cáncer urotelial, melanoma, adenocarcinoma gástrico, hepatocarcinoma y cáncer de pulmón de célula no pequeña (Loesch et al., 2020). Existen varios estudios que apoyan la teoría de que ARID2 se comporta generalmente como supresor tumoral en distintos contextos. En melanoma, aproximadamente el 10% de las muestras analizadas por Hodis et al. presentaron mutaciones sin sentido posiblemente relacionadas con pérdida de función y disminución de capacidad de unión al ADN para llevar a cabo su función reguladora transcripcional (Hodis et al., 2012). En adenocarcinoma de pulmón, hasta un 5% de los pacientes presentaron una ausencia o forma truncada de ARID2 (Manceau et al., 2013).

## **1.5 ARN no codificantes**

Los ARN no codificantes son moléculas no traducidas a proteínas que desempeñan un papel importante en el control de la expresión génica y proteica. Dependiendo de sus funciones dentro de las células, pueden intervenir en tres contextos principales: regulación de la expresión génica (por ejemplo, los micro ARN (miARN) y los ARNlnc), síntesis de proteínas (por ejemplo, los ARN ribosómicos (ARNr)) y síntesis de ARN (por ejemplo, los ARN pequeños nucleares). En el contexto de la regulación epigenética, la desregulación de estos ARN no codificantes puede desencadenar el desarrollo de enfermedades. Entre los mecanismos de regulación más relevantes, destaca el proceso de silenciamiento postranscripcional llevado a cabo por secuencias cortas de ARN de aproximadamente 22 pares de bases conocidos como miARN. Estas moléculas se unen específicamente a sus dianas mediante complementariedad de bases, impidiendo así la traducción de los ARN mensajeros (ARNm). Además, los miARN pueden ser

objeto de modificaciones químicas que podrían afectar a su función en la regulación de la expresión génica (Pathania et al., 2022; Yao et al., 2019).

Otras de las moléculas de más reciente descubrimiento son los ARNlnc, cuya definición y funciones en la regulación de la expresión génica se explorarán en los siguientes apartados.

### 1.5.1 Definición de ARNlnc

Los ARNlnc se definen de manera arbitraria como una clase de moléculas de ARN, con una longitud mayor de 200 nucleótidos, que no son traducidas a proteínas dado que carecen de marcos abiertos de lectura estables (Mattick et al., 2023; Statello et al., 2021). Según el consorcio GENCODE (del inglés, “*ENCYClopedia Of DNA Elements*”), el genoma humano posee más de 16.000 genes que codifican a ARNlnc, aunque en otros proyectos, por ejemplo, en FANTOM5, este número puede verse incrementado hasta 100.000 aproximadamente (Nojima & Proudfoot, 2022). Aunque se han detectado ARNlnc en diversas patologías, se desconoce cuál es la relevancia funcional y biológica de la gran mayoría de ellos (Statello et al., 2021).

### 1.5.2 Síntesis, procesamiento y localización de los ARNlnc

El inicio de la síntesis de ARNlnc suele producirse como consecuencia de la transcripción de PCG o derivarse de regiones accesibles de la cromatina donde interviene el complejo de transcripción de la Polimerasa II. A su vez, estas regiones accesibles de la cromatina suelen estar asociadas tanto con los promotores como los *enhancers* de PCG (Nojima & Proudfoot, 2022). La presencia de bucles R en estas regiones reguladoras puede facilitar la transcripción de los ARNlnc (Tan-Wong et al., 2019). A diferencia de los ARNm, en la iniciación de la transcripción de PCG, se recluta de manera coordinada otros factores necesarios para el procesamiento del ARN, lo que explicaría la baja frecuencia y eficiencia de *splicing* y poliadenilación del extremo 3' que poseen los ARNlnc (Mélé et al., 2017; Schlackow et al., 2017). En cuanto a la terminación de la transcripción de los ARNlnc, se han descubierto diferentes factores que intervienen en este proceso de manera separada o cooperativa: la endonucleasa Integrator, el complejo Restrictor y el Microprocesador DROSHA (Kamieniarz-Gdula & Proudfoot, 2019; Nojima & Proudfoot, 2022).

En comparación con los ARNms, los ARNlnc se encuentran menos conservados evolutivamente y se encuentran compuestos por menos exones de mayor longitud (Hezroni et al., 2015; Quinn et al., 2016). Además, el nivel de ex-

presión génica de los ARNlnc es menor, ubicándose el producto preferentemente en el núcleo celular (Melé et al., 2017). Tras la transcripción, los ARNlnc pueden ser sometidos al *splicing*, a poliadenilación y a la adición de una caperuza de 7-metil guanosina. No obstante, existen otros que no son modificados tras este proceso, por ejemplo, aquellos expresados a partir de los promotores de la Polimerasa I o Polimerasa III (Gingeras, 2007). A diferencia de los ARNm, la gran mayoría de los ARNlnc no son poliadenilados en el extremo 3' siendo más inestables y con mayor probabilidad de degradación por el complejo exosoma (Schlackow et al., 2017). Además, las señales internas de *splicing* débiles presentes en su secuencia y la larga distancia entre el sitio de corte y el punto de ramificación disminuyen la eficiencia de este procesamiento (C.-J. Guo et al., 2020; Melé et al., 2017). La alta concentración de ARNlnc en el núcleo no solo es debido a este procesamiento post-transcripcional menos eficiente, sino también a la presencia de modificaciones epigenéticas represoras de histonas en sus promotores, su posible asociación específica a la cromatina (Nojima et al., 2018), la presencia de señales de poliadenilación alternativas al extremo 3' u otros motivos específicos de localización nuclear (Statello et al., 2021; Xiang et al., 2014).

Existe una fracción de ARNlnc que es sometida a exportación al citoplasma para asociarse con orgánulos específicos o proteínas de unión a ARN. De esta manera, pueden ejercer distintas funciones esenciales fuera del núcleo, por ejemplo, controlar la estabilidad de los ARNm, intervenir en rutas específicas de señalización o regular la traducción de proteínas (Statello et al., 2021). Por último, de manera externa a la célula, se ha descubierto recientemente la presencia de ARNlnc en exosomas obtenidos en muestra de sangre humana (S. Li et al., 2018; Statello et al., 2018),

### 1.5.3 Clasificación de los ARNlnc

Debido a la gran heterogeneidad de los ARNlnc, se pueden clasificar según distintos criterios, basados principalmente en su posición en el genoma o en su función (Cruz-Miranda et al., 2019):

- Según su posición genómica y su orientación en relación con los PCG vecinos, los ARNlnc se pueden clasificar como ARNlnc sentido, ARNlnc antisentido, ARNlnc intrónicos y ARNlnc intergénicos. Los ARNlnc sentido son aquellos que se superponen con uno o más exones de otro gen codificante y son transcritos en el mismo sentido que este. Por otro lado, los ARNlnc antisentido son aquellos que se superponen al gen codificante en la cadena complementaria y son transcritos en el sentido contrario. Los ARNlnc intrónicos se encuentran ubicados en regiones intrónicas de genes codificantes y su transcripción con frecuencia es

bidireccional. Por último, los ARNlnc intergénicos se generan a partir de secuencias genómicas entre dos genes codificantes. Se piensa que los dos últimos tipos mencionados, los ARNlnc intrónicos e intergénicos, son regulados mayoritariamente a través de mecanismos de transcripción y modificación de cadena poli-A específicos (L. Ma et al., 2013; Yousefi et al., 2020).

- Según la distancia a la que ejerza su regulación o su mecanismo de acción, los ARNlnc puede actuar en *cis* o en *trans*. Aquellos ARNlnc que actúan en *cis* no suelen abandonar la cromatina para ejercer su acción de regulación de la expresión génica en genes cercanos a su sitio de transcripción. Por el contrario, los ARNlnc *trans* regulan la transcripción de genes lejanos a su origen, pueden modificar la organización y la estructura nuclear, e incluso pueden interaccionar con otras moléculas de ARN y proteínas para modular sus comportamientos en el citosol (Kopp & Mendell, 2018).
- Según el efecto genómico funcional asociado, que puede verse relacionado con promotores, *enhancers*, elementos transposones y su propia estructura. Los ARNlnc pueden actuar como estructuras de soporte para el ensamblaje de complejos ribonucleoproteicos o de otras estructuras nucleares que puedan modular la transcripción de genes. Existen ARNlnc que puede funcionar como ribo-activadores o ribo-represores modificando la función de diversas proteínas. Por otro lado, los ARN endógenos competitivos, también denominados esponjas de ARNlnc, actúan a nivel postranscripcional atrapando los miARN (Jarroux et al., 2017).

#### 1.5.4 Funciones biológicas de los ARNlnc

Atendiendo al último sistema de clasificación mencionado en el anterior apartado, los ARNlnc pueden intervenir en la regulación de una amplia variedad de funciones moleculares celulares a través de distintos mecanismos de acción tanto a nivel nuclear como citoplasmático (Jarroux et al., 2017). Las interacciones generadas de estas moléculas con su entorno permiten regular la transcripción y la arquitectura de la cromatina, el procesamiento y la estabilidad del ARN, el splicing, o incluso la traducción de proteínas y su localización. (Mattick et al., 2023) [Figura 8].

### 1.5.4.1 Regulación transcripcional

En la gran mayoría de los casos, la función principal de los ARNlnc se basa en su papel principal como regulador transcripcional. Los ARNlnc que intervienen en la regulación transcripcional pueden ser divididos en dos subtipos principales de acuerdo su mecanismo de acción: interferencia o potenciación transcripcional, o regulación estructural de la cromatina. Los ARNlnc pueden regular directamente la expresión génica mediante su interacción con las secuencias *enhancers*. Específicamente, la expresión génica de los ARNlnc *enhancer*, derivados de la transcripción bidireccional de regiones *enhancer* del genoma, se correlaciona positivamente con los niveles de ARNm de los PCG vecinos (Watabe, 2015). Por ejemplo, los ARNlnc *enhancers* producidos a través de la regulación del receptor de andrógenos podrían estar implicados en la regulación transcripcional del desarrollo, la diferenciación celular y el progreso de la enfermedad (D. Wang et al., 2011). De manera adicional, se piensa que los mismos ARNlnc pueden actuar como *enhancers*, a través del reclutamiento de factores de transcripción o el desplazamiento de otras moléculas represoras de la expresión en el promotor de un gen (Ørom et al., 2010).

Por otro lado, de acuerdo con su segundo mecanismo de acción mencionado anteriormente, los ARNlnc pueden jugar un papel crítico en la remodelación de la estructura de la cromatina regulando así indirectamente la expresión génica. Los ARNlnc, tales como *HOTAIR* o *ANRIL*, se encuentran estrechamente relacionados con los complejos remodeladores de la cromatina y los modificadores epigenéticos (Hanly et al., 2018; Watabe, 2015). En un estudio sobre 3.300 ARNlnc humanos, se concluyó que hasta aproximadamente el 20% de ARNlnc estudiados interactuaban con los complejos nucleoproteicos PRCs (del inglés, “*Polycomb Repressive Complex*”), que promueven el silenciamiento de los genes a través de modificaciones de la estructura de la cromatina (Khalil et al., 2009; Long et al., 2017). En este contexto, se piensa que los ARNlnc intervienen de dos maneras distintas: pueden generar un andamiaje estable para unirse o facilitar la formación de un complejo proteico; o puede proporcionar especificidad al sistema guiando al complejo remodelador de la cromatina a una secuencia específica de ADN (Dykes & Emanuelli, 2017).

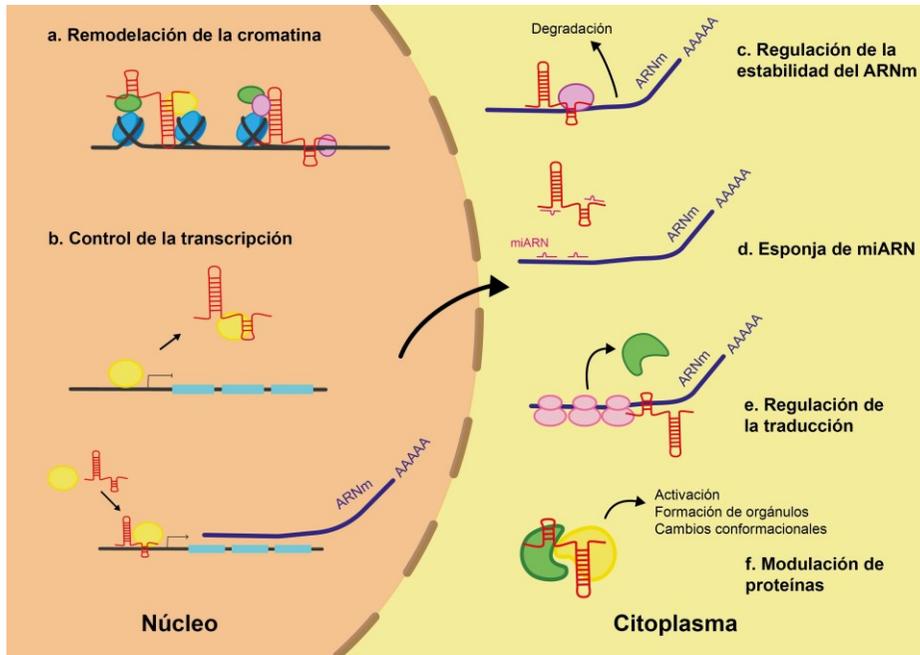
### 1.5.4.2 Regulación postranscripcional.

Los ARNlnc también se encuentran involucrados en la regulación postranscripcional de los ARNm a través de distintos mecanismos. Aunque los ARNlnc pueden servir como fuente de síntesis de miARN, también pueden secuestrar estos ARN no codificantes cortos que inhiben la expresión génica. Muchos ARNlnc contienen en su estructura sitios de unión de miARN, quedando

secuestrados y limitando su concentración intracelular y su función regulatoria. Así, estos ARNlnc, conocidos en la bibliografía como ARNlnc endógenos competitivos, actúan como reguladores negativos de los miARN y, por tanto, como reguladores positivos de la expresión génica (Dykes & Emanuelli, 2017; Salmena et al., 2011). Por ejemplo, *XIST* es un ARNlnc que inhibe la invasión tumoral en el cáncer de mama mediante el secuestro del miARN miR-155 (Zheng et al., 2018).

Por otro lado, los ARNlnc mantienen una red de interacciones con los ARNm en sus distintos niveles de procesamiento. Por ello, se pueden encontrar varios ejemplos de ARNlnc caracterizados en la bibliografía dentro de las estructuras subnucleares “*paraspeckles*”. Estos orgánulos se encuentran compuestos principalmente por hasta aproximadamente 40 proteínas (la mayoría de ellas proteínas de unión a ARN) y el ARNlnc *NEAT1*. Aunque la función de esta estructura aún es desconocida, se conoce que participa en la regulación y retención de los ARNm, el procesamiento de miARN y en el secuestro de proteínas (Fox et al., 2018). Además de *NEAT1*, se puede encontrar otros ejemplos de ARNlnc localizados en los “*paraspeckles*”, como puede ser *MALAT1*, implicado en la regulación del *splicing* alternativo de los ARNm (Gutschner et al., 2013).

De manera independiente a la anterior estructura, los ARNlnc pueden inhibir la traducción del ARNm o degradarlo directamente gracias a su capacidad de unión a proteínas o a ácidos nucleicos. Gracias a la presencia de motivos específicos en su secuencia, los ARNlnc pueden unirse a proteínas de unión a ARN para inhibir el proceso de *splicing* alternativo, por ejemplo, el ARNlnc *PNCTR* secuestra a la proteína PTBP1 (Yap et al., 2018). En el caso de degradación de ARNm, los ARNlnc s con elementos Alu, pueden unirse a la región 3'UTR para que intervenga la proteína de degradación de ARNm *STAU1* en su eliminación (Dykes & Emanuelli, 2017; Geisler & Coller, 2013).



**Figura 8.** Mecanismos de regulación de los ARNnc. **a)** Interacción con complejos remodeladores de la cromatina. **b)** Regulación transcripcional de la expresión génica mediante su interacción directa con secuencias regulatorias, o actuando como señuelo para otros factores de transcripción (círculos amarillos). **c)** Regulación de la estabilidad del ARNm. **d)** Control de los niveles de ARNm mediante su interacción con miARN. **e)** Regulación de la traducción de los ARNm. **f)** Control de las funciones de las proteínas. Adaptación de (Gao et al., 2020).

# Capítulo II

## Materiales y Métodos

# CAPÍTULO II: Materiales y Métodos

## 2.1 Análisis estadístico

La normalidad de los datos continuos fue analizada mediante el uso de gráficos de cuantiles-cuantiles y la prueba de Shapiro-Wilk. Para los datos normales, se obtuvo la media y la desviación estándar, y se aplicaron pruebas t de Student de dos colas tras confirmar la igualdad de varianzas. En el caso de que no existiese homocedasticidad, se aplicó la corrección de Welch. Para los datos no normales, se obtuvo la mediana y los cuantiles, y se aplicaron pruebas de suma de rangos de Wilcoxon. En el caso de realizar muchas comparaciones, se utilizó el método de Benjamini-Hochberg para controlar la tasa de descubrimiento falso (FDR, del inglés “*False Discovery Rate*”) obteniendo p-valores ajustados (representados en el texto como FDR). Por último, para los análisis de correlación específicos de la expresión génica, se empleó la prueba de correlación de Spearman.

## 2.2 Programas, herramientas online y bases de datos

Los análisis bioinformáticos realizados fueron ejecutados utilizando el lenguaje de programación R (versión 4.1) y paquetes de funciones derivados de Bioconductor (versión 3.14). Los análisis que requerían altos recursos computacionales fueron llevados a cabo en el clúster de supercomputación de alto rendimiento Alhambra (<http://alhambra.ugr.es/>) utilizando scripts codificados en bash.

Para realizar los análisis correspondientes a los resultados de los Capítulos, se recurrieron a los siguientes recursos bioinformáticos a nivel local y externo especificados en la **Tabla 1**.

**Tabla 1.** Recursos externos y *softwares*. El valor “-“ indica que se utilizó la versión más actualizada o distintas versiones según la fecha de uso.

Recurso	Versión	URL/Referencia bibliográfica
Agilent Feature Extraction Software	11.0.1.1	<a href="https://www.agilent.com/en/product/mirna-microarray-platform/mirna-microarray-software/feature-extraction-software-228496">https://www.agilent.com/en/product/mirna-microarray-platform/mirna-microarray-software/feature-extraction-software-228496</a>

DepMap	-	<a href="https://depmap.org/portal/ccle/">https://depmap.org/portal/ccle/</a>
Enrichr	-	(E. Y. Chen et al., 2013; Kuleshov et al., 2016)
Ensembl	-	<a href="https://www.ensembl.org/">https://www.ensembl.org/</a>
FASTQC	0.12.1	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
FlowJo	10.4	<a href="https://www.flowjo.com/">https://www.flowjo.com/</a>
GDC	-	<a href="https://gdc.cancer.gov/">https://gdc.cancer.gov/</a>
GeneHancer	-	(Fishilevich et al., 2017)
Genoma de ratón (mm39)	VM32	<a href="https://www.gencodegenes.org/mouse/release_M32.html">https://www.gencodegenes.org/mouse/release_M32.html</a>
IGV	2.16.1	(Robinson et al., 2011)
MSigDB	-	<a href="https://www.gsea-msigdb.org/gsea/msigdb">https://www.gsea-msigdb.org/gsea/msigdb</a>
Qualimap	2.2.1	(García-Alcalde et al., 2012)
R	4.1	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
REVIGO	-	(Supek et al., 2011)
Samtools	1.17	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
STAR	v2.7.10b	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
UCSC Genome Browser	-	<a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a>

## 2.3 Microarray de ARNInc

### 2.3.1 Muestras de pacientes con LLA-B y sanas

Se obtuvieron 42 muestras de médula ósea de pacientes pediátricos con LLA-B y 4 muestras de médula ósea de donantes sanos. Todos los pacientes proporcionaron su consentimiento informado por escrito para el uso de sus muestras y datos clínicos con fines de investigación de acuerdo con la Declaración de Helsinki. Los tejidos de médula ósea se congelaron al momento de la adquisición y se almacenaron a  $-80\text{ }^{\circ}\text{C}$ . Todos los procedimientos, manipulación y almacenamiento de las muestras fueron los mismos para los pacientes pediátricos con LLA-B y las muestras controles. La información clínica fue recopilada a partir de informes médicos, pudiéndose consultar en la información suplementaria de (Cuadros et al., 2019). La mayoría de los casos fueron tratados utilizando protocolos de terapia similares basados en los españoles PETHEMA y SEHOP. El estudio fue aprobado por el Comité de Ética del Hospital Universitario Virgen de las Nieves (Granada, España).

### 2.3.2 Diseño del microarray y predicción de la regulación de ARNlnc

El *LncPath*<sup>TM</sup> *Human Cancer Array* fue diseñado para medir la expresión de 2.829 ARNlnc y 1.906 PCG. La lista de ARNlnc fue extraída de distintas fuentes de información: GENCODE, RefSeq, GenBank, UCSC Genome Browser y genes descritos por el estudio de Cabili et al. (Cabili et al., 2011). Por otro lado, los PCG fueron seleccionados en base a la predicción de regulación ejercida por los ARNlnc.

La lista de ARNlnc seleccionados previamente para el análisis se clasificaron en función de su interacción predicha con los PCG en tres grupos:

- ARNlnc vecinos, considerados como ARNlnc que se encuentran a menos de 3 kb de genes clave de rutas de cáncer. Estos ARNlnc podrían modular la expresión de PCG vecinos.
- ARN endógenos competitivos (ARNec), que comparten elementos de respuesta de miARN con los ARNm de genes claves de rutas y, por lo tanto, podrían evitar que estos ARNm sean degradados. Estos ARNec podrían afectar a la expresión de los PCG que son modulados por los miARN relacionados.
- ARNlnc similares a *enhancers*, que se localizan dentro de los 300 kb de gene clave de rutas de cáncer y podrían actuar como *enhancers* según publicaciones científicas.

### 2.3.3 Etiquetado de ARN e hibridación

Para la amplificación del ARN, se utilizó el kit Epicentre's TargetAmp 1-Round aRNA Amplification (Illumina, #TAB1R80524). Tanto la preparación de la muestra como el proceso de hibridación fueron realizados de acuerdo con el protocolo estándar del fabricante. El ARN total de cada muestra fue amplificado y transcrito a ARNc siguiendo el protocolo Agilent's Quick Amp Labeling (v5.7, Agilent Technologies). Posteriormente, los ARNc etiquetados se hibridaron en el *LncPath*<sup>TM</sup> *Human Cancer Array* (8x15K, Arraystar). Tras el lavado establecido en el protocolo, los arrays fueron escaneados en el Agilent Scanner G2505C. Las

imágenes obtenidas fueron analizadas mediante el uso del programa Agilent Feature Extraction (v11.0.1.1).

### 2.3.4 Análisis de datos de expresión génica

Los valores de expresión génica crudos y normalizados obtenidos para cada una de las muestras en el análisis del *microarray* están disponibles públicamente en el repositorio GEO (del inglés, “*Gene Expression Omnibus*”) GSE128254. Para realizar el análisis de expresión, se utilizó el paquete de R *limma* (v3.50). En primer lugar, se realizó una normalización de cuantiles en los datos crudos. Se realizó un filtrado de los valores de intensidad bajos, manteniendo aquellos ARNlnc/ARNm donde al menos 4 de las 46 muestras estuvieran etiquetadas con “P” o “M”, es decir, con valores disponibles en todas las dianas. Los ARNlnc/ARNm diferencialmente expresados entre las muestras tumorales y normales se obtuvieron utilizando como umbrales un cambio relativo absoluto (FC, del inglés “*Fold Change*”) mayor a 1,5 y un FDR menor de 0,05. Posteriormente, se realizó un agrupamiento jerárquico basado en el coeficiente de correlación de Spearman para identificar patrones de expresión de ARNlnc y ARNm entre las muestras tumorales y las muestras sanas.

### 2.3.5 Análisis de ontología de genes

Se realizó un análisis de ontología de genes (GO, del inglés “*Gene Ontology*”) para identificar vías biológicas, funciones moleculares o componentes celulares significativamente enriquecidos relacionados con los ARNlnc diferencialmente expresados identificados en el análisis del *microarray* entre muestras tumorales y normales ((FC absoluto) > 1,5, FDR < 0,05). Tal como se mencionó anteriormente, los PCG asociados a los pares de ARNlnc/ARNm diferencialmente expresados fueron predichos por el fabricante del *microarray* en función de su proximidad genómica, predicción de uniones de miARN y la literatura científica, según se describe en la sección “2.3.2 Diseño del *microarray* y predicción de la regulación de ARNlnc”. Usando estas predicciones, se realizó el análisis de GO de la siguiente manera:

- 1) Se detectaron aquellos ARNlnc regulados al alza o a la baja en la comparativa entre muestras tumorales de LLA-B pediátrica y médulas óseas sanas ((FC absoluto) > 1,5, FDR < 0,05). Estos ARNlnc regulados al alza y a la baja fueron analizados de manera separada para los siguientes pasos.

- 2) Usando las predicciones de los pares ARNlnc/ARNm, se seleccionaron aquellos ARNm de PCG predichos que estaban asociados a los ARNlnc seleccionados del anterior punto.
- 3) Se utilizó la lista de ARNm del anterior punto en la base de datos Enrichr (E. Y. Chen et al., 2013; Kuleshov et al., 2016) para encontrar términos GO significativamente enriquecidos.
- 4) Con el fin de eliminar redundancias en los términos GO, se utilizó ReviGO (Supek et al., 2011) con los parámetros por defecto y proporcionando los valores FDR para cada término GO.

## 2.4 Datos externos: TCGA y CCLE.

Los datos descargados de pacientes y líneas celulares fueron utilizados para comparar la expresión de los genes de interés entre subtipos celulares B y T de la LLA y los pacientes sanos, para realizar análisis de correlación génica y para examinar la supervivencia de los pacientes.

### 2.4.1 Análisis de datos de expresión de pacientes

#### 2.4.1.1 Cohorte de pacientes con LLA pediátricos

Los datos de expresión génica de los pacientes con LLA se obtuvo a partir del portal Genomic Data Commons (GDC) a través del proyecto TARGET (del inglés, “*Therapeutically Applicable Research To Generate Effective Treatments*”) (descarga: 06-06-2020; N = 506). Los datos se descargaron utilizando el paquete de R TCGABiolinks (v2.22.4) a través de la función `GDCquery` con los siguientes parámetros:

```
GDCquery(project = list("TARGET-ALL-P1", "TARGET-ALL-P2",  
"TARGET-ALL-P3"), data.category = "Transcriptome Profiling",  
data.type = "Gene Expression Quantification", experimental.strategy =  
"RNA-Seq", workflow.type = "STAR - Counts", sample.type = "Primary Blood  
Derived Cancer - Bone Marrow")
```

Para obtener la cohorte de pacientes pediátricos con LLA, se utilizaron los datos clínicos obtenidos a partir del portal GDC. Para realizar los posteriores análisis, se descartaron los pacientes que tenían más de 14,99 años. En el caso de muestras procedentes del mismo paciente, se eliminaron aquellas con el tamaño

de biblioteca más pequeño. Por último, se seleccionaron las muestras que tenían información disponible en los datos clínicos sobre el subtipo celular de LLA del paciente: B (N = 120) y T (N = 200). En este caso, el conjunto de datos de subtipo celular B consistía en muestras de subtipo pre-B (112/120, 93,33%) y LLA-B (8/120, 6,67%), mientras que todas las muestras de subtipo T eran de LLA-T.

### 2.4.1.2 Cohorte de pacientes sanos

Para la obtención de los datos de expresión génica de pacientes sanos procedentes del proyecto TARGET (descarga: 26-07-2023; N = 88), que estuvieron disponibles durante el desarrollo de la investigación detallada en el Capítulo IV, se utilizaron los siguientes parámetros:

```
GDCquery(project = "TARGET-AML", data.category =
"Transcriptome Profiling", data.type = "Gene Expression
Quantification", sample.type = "Bone Marrow Normal")
```

Debido a que no existían datos clínicos disponibles sobre estas muestras para saber si procedían de un paciente curado o tratado, se descartaron aquellas que tenían una muestra pareada tumoral y poseían, según la nomenclatura de la Oficina de Genómica del Cáncer (OCG, del inglés "*Office of Cancer Genomics*") el código tumoral 20 asignado que corresponde a LMA. En este punto, se asumió que las muestras obtenidas (N = 50) procedían de personas que no habían padecido LMA.

### 2.4.1.3 Normalización de los datos de expresión

En cuanto al tratamiento de los datos de expresión génica de las cohortes obtenidas, las lecturas crudas de cada gen fueron normalizadas utilizando el método TMM (del inglés, "*Trimmed Mean of M Values*") con el fin de ajustar por el tamaño de librería de las muestras. Posteriormente, se obtuvieron los valores  $\log_2$  de las lecturas por millón (CPM, del inglés, "*Counts per Million*") utilizando el paquete de R `edgeR` (v3.30.3) sumando una unidad (+1) con el fin de evitar  $\log_2$  de cero. El código genérico utilizado para obtener estos valores fue:

$$\log_2(\text{cpm}(\text{tmm}) + 1),$$

donde "cpm" representa una función del paquete de R `edgeR` para computar las lecturas por millón, y "tmm" la matriz de lecturas normalizadas.

#### 2.4.1.4 Análisis de supervivencia

Los niveles de expresión de los genes estudiados fueron utilizados para analizar su correlación con la supervivencia global de los pacientes con LLA pediátricos, considerando las muertes como eventos. Los subgrupos de los pacientes a comparar se obtuvieron dicotomizando la expresión génica del gen analizado en “Bajo” o “Alto” según si el nivel se encontraba por debajo o por encima del valor umbral de la mediana de expresión de la cohorte correspondiente. En el caso específico de la cohorte de pacientes con LLA-T pediátrica, para la expresión de los genes AL133346.1 y CCN2, los subgrupos “Bajo” y “Alto” se obtuvieron según si los valores de expresión eran nulos o no.

La asociación entre las características clínicas y los subgrupos de expresión génica se determinó mediante pruebas exactas de Fisher. Se realizaron curvas de supervivencia global Kaplan-Meier comparando los subgrupos de pacientes “Alto” y “Bajo” de los genes de interés mediante los paquetes de R `survival` (v3.5-7) y `survminer` (v0.4.9). La significancia estadística del resultado se obtuvo mediante la prueba de rango logarítmico univariante. Adicionalmente, se estudió el efecto de las diferentes covariables clínicas disponibles en los datos descargados en la supervivencia global de los pacientes mediante el uso de la regresión de Cox univariante. Aquellas covariables clínicas que mostraron  $p$ -valor  $< 0,2$  en el análisis univariante se utilizaron para realizar la regresión de Cox multivariante.

#### 2.4.2 Análisis de datos de expresión de líneas celulares

Los datos de expresión génica normalizados por TPM (del inglés, “*Transcripts Per Million*”) se descargaron directamente de la base de datos CCLE (del inglés, “*Cancer Cell Line Encyclopedia*”), contenida actualmente en Depmap (<https://depmap.org/portal/>) (descarga: 29-09-2019; N = 1019). Utilizando la información que describe las características de las líneas celulares procedente del portal DepMap, se seleccionaron aquellas líneas celulares que estaban catalogadas como LLA-B (N = 14) y LLA-T (N = 16).

Más tarde, para el análisis realizado en el Capítulo IV, se realizó una nueva descarga de los datos de expresión actualizados de líneas celulares procedentes del portal Depmap (versión 24-06-2023; N = 1450). En este caso, se pudo ampliar la población de líneas celulares de LLA-B a 24, mientras que la de LLA-T se redujo a 14.

En ambos casos para los posteriores análisis, los datos de expresión génica normalizados por TPM obtenidos se transformaron a escala  $\log_2$  sumando una unidad (+1) con el fin de evitar  $\log_2$  de cero.

## 2.5 Secuenciación ARN-Seq

### 2.5.1 Genotipo de ratones

Los experimentos realizados con el organismo modelo *Mus musculus* se realizaron de manera controlada respecto a la edad y al sexo. Todos los ratones *Arid1a<sup>ff</sup>Arid1b<sup>ff</sup>*, con y sin el sistema UBC-CreER, poseían el fondo genético de una mezcla de las cepas C57BL/6 y 129P2/OlaHsd. Para obtener el modelo doble knock-out (DKO), basado en la eliminación del exón 9 de *Arid1a* y el exón 5 de *Arid1b*, se inyectó 75 mg de tamoxifeno (ThermoFisher Scientific, #J63509) por kg de peso intraperitonealmente a todos los ratones de 3 meses de edad durante 5 días consecutivos. En el caso de los fibroblastos embrionarios derivados de estos modelos de ratón, se trataron con una disolución 0,5  $\mu\text{M}$  de tamoxifeno durante 48 horas para producir la delección.

### 2.5.2 Muestras

Las muestras secuenciadas en el ARN-Seq proceden de cultivos primarios de fibroblastos embrionarios de ratón (MEF, del inglés “*Mouse Embryonic Fibroblasts*”) y tejidos de hígado e intestino delgado extraídos de *Mus musculus*. En el caso de las líneas celulares, se secuenciaron 3 muestras controles y 3 muestras tratadas con tamoxifeno, mientras que en el caso de los tejidos se obtuvieron 5 muestras de cada grupo para cada uno de ellos. El método de extracción del ARNm se encuentra especificado en la sección “2.6.1 Extracción y análisis de ARNm”.

### 2.5.3 Protocolo de secuenciación

La preparación de las librerías de los protocolos especificados a continuación fue realizada por la Unidad de Genómica de GENYO (Centro Pfizer-Universidad de Granada-Junta de Andalucía).

Para las muestras procedentes de los cultivos primarios de MEF, se utilizó el protocolo de secuenciación TruSeq Stranded Total RNA (Illumina). Se obtuvieron tres réplicas biológicas en cada condición experimental. Las lecturas obtenidas con este protocolo son *paired-end* con una longitud de 75 pares de bases.

En el caso de las muestras de hígado y yeyuno del animal de experimentación *Mus musculus*, se utilizó el protocolo de secuenciación Illumina Stranded mRNA Prep (Illumina). Se obtuvieron cinco réplicas biológicas en cada condición experimental. Al igual que las muestras procedentes de líneas celulares, las lecturas obtenidas son *paired-end* con una longitud de 75 pares de bases.

### 2.5.4 Procesamiento de datos de ARN-Seq

En primer lugar, se analizó la calidad de los archivos de secuenciación FASTQ derivados de las muestras mediante la herramienta FASTQC (v0.12.1). Posteriormente, las lecturas fueron alineadas en el genoma de referencia de ratón mm39 (VM32) previamente indexado, utilizando la herramienta STAR (v2.7.10b) con las siguientes opciones:

```
--twopassMode Basic, --twopass1readsN -1, outSAMtype BAM
SortedByCoordinate, --sjdbOverhang 74, --outSAMunmapped
Within, --clip3pAdapterSeq CTGTCTCTTATACACATCT CTGTCTCTTA-
TACACATCT, --clip3pAdapterMMP 0.1 0.1, --quantMode Gene-
Counts
```

En este caso, se utilizó la opción “`--quantMode GeneCounts`” disponible de la herramienta STAR para obtener los conteos de lecturas por gen necesarios para realizar los análisis de expresión diferencial. Por último, se realizó un análisis de calidad de los archivos BAM obtenidos con la herramienta Qualimap (v.2.2.1).

### 2.5.5 Análisis de expresión diferencial

Para realizar el análisis de expresión diferencial, se utilizó el paquete de R DESeq2 (v1.34.0). En este punto, se consideraron como genes expresados diferencialmente de manera significativa aquellos que presentaban un valor absoluto de FC > 1,5 y un FDR < 0,05 entre los grupos de muestras comparados. Para los posteriores análisis, los valores de los log<sub>2</sub>FC fueron ajustados utilizando el método “`ashr`” (Stephens, 2016), preservando los p-valores y FDR previamente obtenidos.

Para la visualización de los datos y la exploración de la similitud de las muestras, se realizó la regularización del logaritmo mediante la función `rlog` con el fin de estabilizar la varianza de la expresión de los genes. Los nuevos valores fueron utilizados para realizar un análisis de componentes principales (PCA, del inglés, “*Principal Component Analysis*”). De manera excepcional, antes de realizar el PCA de las muestras procedentes de MEF, se eliminó el efec-

to *batch* mediante el uso de la función `removeBatchEffect` del paquete `limma` (v3.50.3).

## 2.5.6 Análisis de enriquecimiento en conjuntos de genes

En primer lugar, se obtuvieron los identificadores y símbolos de los genes con el uso de la anotación contenida en Ensembl (v109) para *Mus musculus* mediante el paquete de R `biomaRt` (v2.50.3). Además, se descargaron las rutas biológicas para este organismo modelo a partir de la base de datos MSigDB, obteniéndose las versiones 7.1 de los conjuntos: Hallmarks, Reactome, KEGG, BioCarta. A este grupo de rutas, se añadieron los términos GO descargados mediante el paquete de R `org.Mm.eg.db` (v3.14.0).

Para realizar el GSEA, se empleó el paquete de R `fgsea` (v1.20.0) utilizando los valores de los  $\log_2FC$  ajustados obtenidos en el anterior paso para ranquear la lista de genes. Por último, antes de obtener la lista de rutas biológicas enriquecidas en el ARN-Seq, se empleó la función `collapsePathways` para optimizar la selección de rutas independientes.

## 2.6. Métodos de biología molecular

### 2.6.1 Extracción y análisis de ARNm

La extracción total de ARN se realizó a partir de *pellets* celulares utilizando el reactivo TRIzol (Sigma-Aldrich, #T9424), siguiendo el protocolo estándar del fabricante. La concentración y la calidad de las muestras de ARN obtenidas fueron medidas mediante el uso de un espectrofotómetro Nanodrop 2000 (ThermoFisher Scientific, ND-2000). En el caso de las muestras derivadas de tejido destinadas a secuenciación ARN-Seq, con el fin de optimizar y mejorar la calidad del ARN, se realizó la extracción mediante el uso del kit RNeasy (QUIAGEN, #74104). La concentración y calidad de las muestras de ARN para secuenciación fueron analizadas en el Bioanalyzer 2100 (Agilent Technologies) para asegurar la calidad mediante el análisis del valor RIN (del inglés, “RNA Integrity Number”). Las muestras de ARN total obtenidas fueron almacenadas a -80 °C para preservar su integridad.

## 2.6.2 RT-qPCR

Para la síntesis de ADN complementario (ADNc), se usaron 1-2  $\mu\text{g}$  de ARN extraído utilizando el kit RevertAid RT (ThermoFisher Scientific, #K1691) mediante el uso de hexámeros aleatorios. Previamente, siguiendo el protocolo estándar del fabricante, se eliminó el posible ADN genómico presente en las muestras mediante una digestión enzimática con DNase I (Invitrogen, #18068015). Las reacciones de qPCR fueron llevadas a cabo en placas de 96 pocillos siguiendo el protocolo de KAPA SYBR® FAST (Merck, #KK46) en el QuantStudio™ 3 Real-Time PCR System (ThermoFisher Scientific, #A28567). Todas las muestras fueron medidas por triplicado usándose 50-100 ng de ADNc molde por pocillo. El programa de qPCR utilizado fue: 2 minutos a 50 °C, 10 minutos a 95 °C, y posteriormente 15 segundos a 95 °C y 1 minuto a 60 °C durante 40 ciclos. Los cebadores empleados en la cuantificación de los genes de estudio se encuentran especificados en la **Tabla 2**.

Por último, el cálculo de la expresión relativa del gen de interés se realizó con el método  $\Delta\Delta\text{Ct}$ . La normalización de la expresión entre muestras se obtuvo con el gen constitutivo GAPDH.

**Tabla 2.** Lista de cebadores para la cuantificación de la expresión génica mediante qPCR.

Nombre	Secuencia (5'-3')	Tm (°C)
Ascl2_Mm_qPCR_Fw	GAGACAAGCGTCTCCGGAAT	65,6
Ascl2_Mm_qPCR_Rv	GCCTGGAAGCCCAAGTTTAC	64,6
Gapdh_Mm_Fw	GGGTGTGAACCACGAGAAAT	57,7
Gapdh_Mm_Rv	CCTTCCACAATGCCAAAGTT	55,8
Lgr5_Mm_qPCR_Fw	TCCTAGAAGAGTTACGTCTTGCT	59,2
Lgr5_Mm_qPCR_Rv	CCTTGGGAATGTGTGTCAAAGC	59,2

## 2.6.3 Extracción, separación y cuantificación de proteínas

En cuanto a la extracción total de proteínas, los precipitados celulares obtenidos en los distintos experimentos fueron resuspendidos en el *buffer* de lisis RIPA (150 mM NaCl, 1% NP-40, 0,5% desoxicolato de sodio, 0,1% dodecilsulfato sódico (SDS, del inglés “*Sodium Dodecyl Sulfate*”), 50 mM Tris-HCl a pH 7,5) junto a inhibidores de fosfatasa y proteasas: 7 mM ortovanadato de sodio, fluoruro de fenilmetilsulfonilo (PMSF, del inglés “*Phenylmethylsulfonyl Fluoride*”) y Pierce™ Protease Inhibitor Minitablets (ThermoFisher Scientific,

#A32955). Una vez añadida esta disolución según el tamaño del precipitado celular obtenido (20-50  $\mu\text{L}$ ), se incubaron las muestras a 4 °C durante 20 minutos sometidas a agitación en Vórtex de manera intermitente. Posteriormente, las muestras fueron centrifugadas a 16.000 g a 4 °C durante 15 minutos. Los sobrenadantes obtenidos se guardaron para cuantificar la concentración proteica mediante el método de Bradford (VWR, #A6932) a 600 nm en el GloMax® Microplate Reader (Promega, #GM3000).

En cuanto a la separación y detección de proteínas mediante Western blot, las muestras de proteínas (20-50  $\mu\text{g}$ ) fueron sometidas electroforesis en gel de poliacrilamida con SDS (SDS-PAGE, del inglés “*SDS-Polyacrylamide Gel Electrophoresis*”) con el fin de segregar las proteínas en función de su peso molecular en condiciones desnaturalizantes. Una vez separadas, las proteínas fueron transferidas a una membrana de polivinilideno difluoruro (PVDF, del inglés “**Polyvinylidene Difluoride**”) (VWR, #10600021) previamente activada con metanol, para la detección y cuantificación de las muestras. Las membranas de PVDF fueron incubadas durante 1 hora con una solución de bloqueo, que contenía: solución salina tamponada con fosfato (PBS, del inglés: “*Phosphate-Buffered Saline*”), 5% leche desnatada en polvo y 0,1% Tween-20. Posteriormente, las membranas fueron incubadas con los anticuerpos primarios específicos de la proteína de interés durante toda la noche a 4 °C en agitación constante. Tras la primera incubación, las membranas fueron lavadas tres veces durante 10 minutos en una disolución de PBS con 0,1% de Tween-20. Por último, las membranas se incubaron con anticuerpos secundarios marcados con peroxidasa de rábano (HRP, del inglés “*Horseradish Peroxidase*”) [Tabla 3].

Para la visualización de los resultados, las bandas específicas de las proteínas de interés fueron reveladas mediante el uso de los reactivos Clarity™ Western ECL Substrate (BioRad, #170-5060) o SuperSignal™ West Femto (ThermoFisher Scientific, #34095). Las imágenes del revelado fueron obtenidas en el sistema de imagen ImageQuant™ LAS-4000 (GE Healthcare). La cuantificación de la expresión proteica visualizada en el revelado se realizó mediante densitometría en el software ImageJ (v1.53t). En este caso, la expresión de la proteína de interés se normalizó con la señal obtenida en el control de carga correspondiente ( $\alpha$ -tubulina).

**Tabla 3.** Lista de anticuerpos primarios y secundarios.

Proteína	Referencia Comercial	Dilución
ARID2	Santa Cruz (#sc-166117)	1:500
α-TUBULINA	Santa Cruz (#sc-23948)	1:10.000
HRP-Mouse IgG	Dako (#P0447)	1:5.000
HRP-Rabbit IgG	Dako (#P0448)	1:2.000

## 2.6.4 Cultivos celulares

Las líneas celulares de LLA-B utilizadas fueron: RS4;11, y SUP-B15 [Tabla 4]. Las condiciones de medio de cultivo de RS4;11 fueron a 37 °C en una atmósfera de CO<sub>2</sub> al 5% con medio de cultivo RPMI 1640 (Biowest, #L0498-500) suplementado con suero bovino fetal (FBS, del inglés “*Fetal Bovinum Serum*”) (Gibco™ ThermoFisher Scientific, #10270-106) al 10%, 100 U/mL de estreptomicina y penicilina (Merck, #P0781) y L-Glutamina al 1% (Merck, #G7513). En el caso de la línea celular SUP-B15, se cultivó en las mismas condiciones anteriores exceptuando el uso del medio de cultivo IMDM (Gibco™ ThermoFisher Scientific, #12440061) suplementado con FBS al 20% y 0,05 mM 2-mercaptoetanol (ThermoFisher Scientific, #31350010). Ambas líneas celulares utilizadas fueron sometidas a análisis de contaminación por micoplasma utilizando el kit Venor GeM-qEP (Minerva Biolabs, #11-9025).

**Tabla 4.** Características de las líneas celulares de leucemia linfoblástica aguda (LLA) de subtipo B (LLA-B) cultivadas *in vitro*.

Línea Celular	Identificador	Subtipo	Edad	Translocación
RS4;11	ACH-000874	LLA-B	32	t(4;11)(q21;q23) – MLL-AFF1
SUP-B15	ACH-000059	LLA-B	9	t(9;22)(q34;q11.2) – BCR-ABL1

Antes de comenzar a cultivarlas, las líneas celulares se encontraban preservadas en nitrógeno líquido almacenadas en criotubos con FBS y 10% de DMSO (del inglés, “*Dimethyl Sulfoxide*”). Siguiendo las condiciones establecidas en la ATCC (del inglés, “*American Type Culture Collection*”), la renovación del medio de cultivo de RS4;11 y SUP-B15 se realizó cada 2-3 días dependiendo de la densidad celular de la suspensión.

### 2.6.5 Silenciamiento génico mediante ARNsi

La transfección de las líneas celulares RS4;11 y SUP-B15 con ARNsi se realizó con el objetivo de silenciar la expresión génica de ARID2. Además, se utilizó una secuencia *scramble* de ARNsi como control negativo [Tabla 5].

El procedimiento fue realizado en placas de 6 pocillos cultivando 1 millón de células en cada réplica. En primer lugar, tanto los ARNsi de ARID2 como el agente Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen, #13778-150) fueron diluidos de manera separada con medio de suero reducido Opti-MEM™ (Gibco™ ThermoFisher Scientific, # 31985070). Posteriormente, ambas disoluciones se mezclaron en una proporción 1:1 y se incubaron a temperatura ambiente durante 5 minutos con el fin de generar los complejos lipídicos con ARNsi. Finalmente, esta disolución final se añadió a las suspensiones celulares contenidas en la placa de 6 pocillos lentamente de manera gradual.

**Tabla 5.** Lista de secuencias de ARNsi utilizadas en los experimentos de transfección.

Diana	Cadena	Secuencia (5' – 3')
ARID2	Sentido	GAACGGUCUUUCCAAAUCA
	Antisentido	UGAUUUGGAAAGACCGUUC
<i>Scramble</i> (control negativo)	Sentido	UAACGACGCGACGACGUAA
	Antisentido	UUACGUCGUCGCGUCGUUA

Tras los tiempos de espera de transfección en el incubador bajo condiciones estándar de cultivo, se recogieron los *pellets* celulares tratados para cuantificar la expresión proteica de ARID2 o para los ensayos funcionales descritos a continuación.

### 2.6.6 Ensayos de viabilidad celular

El análisis de la viabilidad celular fue llevado a cabo en las líneas celulares RS4;11 y SUP-B15 mediante el uso de resazurina en diferentes puntos temporales: 0 horas, 72 horas (3 días), 120 horas (5 días), tras el tiempo óptimo de transfección de las líneas celulares. El ensayo correspondiente al tiempo 0 fue utilizado para normalizar la siembra realizada en las réplicas.

Inicialmente, se sembraron 2.000 células en placas de 96 pocillos por triplicado, añadiéndose en oscuridad 20 µL de una disolución de resazurina (Sigma-Aldrich, #R7017) 0,1 M diluida 1:200 con medio de cultivo específico para cada

línea celular tratada. Para eliminar cualquier interferencia de fondo que pudiera generar el medio de cultivo en la lectura de la fluorescencia, se añadió la misma solución de medio por triplicado en cada placa. Tras la adición del reactivo resazurina, las células se mantuvieron en condiciones estándar en el incubador durante 4 horas. Por último, se añadió 30  $\mu$ L de SDS 3%, y se cuantificó la fluorescencia a 590 nm en el lector Glomax® Plate Reader (Promega).

### 2.6.7 Ensayos de proliferación celular

La proliferación celular fue medida en la línea celular RS4;11 mediante el uso del kit CellTrace™ Violet para citometría de flujo (ThermoFisher Scientific, #C34557). En primer lugar, se añadió 1  $\mu$ L de la solución de tinción a 5  $\mu$ M por cada mL de células en suspensión. Posteriormente, las muestras fueron incubadas a 37 °C durante 20 minutos en oscuridad con el fin de que la tinción se uniera a las aminas intracelulares de manera permanente. Una vez transcurrida la incubación, los restos de reactivo no conjugado presentes fueron neutralizados con la adición de un volumen de medio de cultivo completo cinco veces superior al volumen inicial. Posteriormente, las muestras fueron incubadas a 37 °C durante 5 minutos y se centrifugaron durante 5 minutos a 1.300 r.p.m. Por último, se eliminaron los sobrenadantes, y los *pellets* celulares obtenidos fueron resuspendidos en PBS en tubos aptos para su posterior medición por citometría de flujo.

La proliferación fue cuantificada en los mismos puntos temporales establecidos en los ensayos de proliferación celular: 0 horas, 72 horas (3 días), 120 horas (5 días). Por tanto, tras la tinción con CellTrace™ Violet llevada a cabo en el día 0, se preservaron células RS4;11 teñidas en crecimiento en condiciones de cultivo estándar para las medidas de los siguientes puntos temporales.

Por último, las muestras celulares obtenidas fueron analizadas en un citómetro de flujo BD FACSVerser (BD Biosciences). Los datos obtenidos de las muestras a partir de 10.000 eventos fueron procesados con el recurso FlowJo™ v10.4 Software (BD Life Sciences). Los restos celulares y los dobletes fueron filtrados mediante los *gates* FSC (del inglés, “*Forward Scatter*”) y SSC (del inglés, “*Side Scatter*”). La tinción con el kit CellTrace™ Violet se evaluó utilizando el canal V-450 para cuantificar las células positivas.

### 2.6.8 Evaluación de parámetros biológicos de ratones

En cuanto a las imágenes de microscopía, los tejidos de los ratones fueron fijados con formalina neutral tamponada al 10% durante toda la noche a temperatura ambiente. Posteriormente, las muestras fueron procesadas con un procesador automático de tejidos e incluidas en cera de parafina. A partir de los bloques generados, se cortaron secciones con un grosor de 3-4  $\mu\text{m}$  y se colocaron flotando en un baño de agua caliente (mantenido a 45 °C) suplementado con una pequeña cantidad de alcohol absoluto. Las secciones cortadas se recogieron en portaobjetos para ser secadas durante algunos minutos.

Las inmunohistoquímicas hematoxilina/eosina-Y, PAS (del inglés, “*Periodic Acid-Schiff*”) y de proliferación con Ki67 fueron realizadas utilizando procedimientos estándar por parte de AtrysHealth SA (Barcelona, España). A continuación, se tomaron imágenes representativas utilizando un microscopio Olympus BX43 Q14 y se evaluó la morfología y las tinciones de los tejidos.

Por último, la cuantificación del colesterol y el glucógeno fue realizada en el Centro de Instrumentación Científica de la Universidad de Granada. En el caso de la medición de glucosa en sangre, se empleó un glucómetro proporcionado por el Departamento de Bioquímica III e Inmunología de la Universidad de Granada.

# **Capítulo III**

## **ARNInc en LLA pediátrica**

# CAPÍTULO III: ARNlnc en LLA pediátrica

El Capítulo III describe los resultados relacionados con el Objetivo 1, basados en el análisis del perfil de expresión de ARNlnc en pacientes pediátricos con LLA en muestras procedentes de nuestro grupo de investigación y de conjuntos de datos externos. Los resultados descritos forman parte del artículo que avala esta tesis doctoral: “*LncRNA-mRNA Co-Expression Analysis Identifies AL133346.1/CCN2 as Biomarkers in Pediatric B-Cell Acute Lymphoblastic Leukemia*” (Cuadros et al., 2020).

## 3.1 Antecedentes

### 3.1.1 ARNlnc reguladores de la hematopoyesis

Tal y como se ha introducido en el Capítulo I, los ARNlnc pueden regular de manera específica la expresión génica de múltiples procesos biológicos a nivel epigenético, transcripcional y postranscripcional. Durante la hematopoyesis, la diferenciación de las HSC se encuentra altamente controlada por la interacción de diferentes cascadas de señalización intracelulares y la regulación de programas transcripcionales. En este contexto, los ARNlnc se encuentran involucrados en la regulación de distintas etapas de la hematopoyesis, el desarrollo y la activación del sistema inmunológico (Cruz-Miranda et al., 2019). Entre los primeros candidatos estudiados, se encuentra el ARNlnc *EGO*, que regula transcripcionalmente la expresión de la proteína granular eosinofílica con el fin de estimular la diferenciación celular de células CD34+ a eosinófilos (Wagner et al., 2007). Actualmente, aunque la función de los ARNlnc durante el proceso hematopoyético aún se encuentra en estudio, cada vez son más los ARNlnc descubiertos que se encuentran desregulados en las neoplasias hematológicas. Así, los ARNlnc pueden ser considerados importantes reguladores emergentes de la diferenciación y la maduración de los distintos linajes del sistema hematopoyético (F.-Y. Wang et al., 2020).

Varias investigaciones recientes han señalado la importancia de los ARNlnc en la regulación de la actividad de los factores de transcripción durante el desarrollo de las células B a partir de las HSC (Attaway et al., 2021). Por ejemplo, el ARNlnc *asPU.1* modula negativamente la expresión del factor de

transcripción PU.1 para regular el desarrollo temprano de las células B (Ebralidze et al., 2008). Por otro lado, diversos estudios han demostrado la especificidad de expresión que muestran los ARNlnc en los distintos estadios de diferenciación de las células B. Petri et al. identificaron hasta 7 redes distintas de ARNlnc asociadas a distintos estadios de desarrollo del linaje de células B (Petri et al., 2015). Sin embargo, debido a la falta actual de estudios sobre la relevancia funcional que poseen estas moléculas no codificantes, se requiere investigar sus dianas específicas moleculares para conocer más en profundidad la regulación que ejercen en las diferentes rutas biológicas celulares.

En el contexto de las neoplasias hematológicas, existe un número limitado de ARNlnc cuya desregulación se ha vinculado directamente al desarrollo de la leucemia. En leucemia mieloide aguda (LMA), se ha observado que la expresión del ARNlnc *HOTAIR* se encuentra desregulada asociándose con tiempos de supervivencia menores en los pacientes (Tang & Hann, 2018). El ARNlnc *BALR-6*, ubicado cerca de genes implicados en el desarrollo de las células B, se encuentra desregulado en LLA-B, incrementando la proliferación celular (Rodríguez-Malavé et al., 2015). Estos ejemplos subrayan el papel crítico que pueden jugar los ARNlnc en el diagnóstico, el pronóstico y estratificación para el tratamiento de la leucemia, proporcionando nuevas vías de investigación en estas enfermedades.

### 3.1.2 ARNlnc *TCL6*: Biomarcador en LLA-B pediátrica

En el grupo de investigación, previo a los resultados presentados en esta tesis doctoral y siendo el doctorando partícipe de este trabajo, se realizó un análisis comparativo de los perfiles de expresión de ARNlnc y sus ARNm asociados en pacientes pediátricos con LLA-B clasificados por la presencia del gen de fusión *ETV6-RUNX1* (Cuadros et al., 2019). La cohorte analizada, que forma parte del estudio de este Capítulo III, se encontraba formada por 42 pacientes pediátricos con LLA-B, donde 24 de ellos eran positivos para la translocación t(12;21), 18 negativos para la misma y 4 muestras de médula ósea de pacientes sanos.

En este estudio, el ARNlnc *TCL6* fue identificado como un nuevo biomarcador diagnóstico de pacientes pediátricos con LLA-B positivos para la translocación t(12;12) *ETV6-RUNX1* tras comparar los perfiles de expresión génica con aquellos sin dicha fusión génica. Además, no se obtuvo un aumento significativo de la cantidad de ARNlnc o ARNm expresados particularmente en los cromosomas 12 o 21, lo que sugería que la translocación afectaba a la expresión génica de manera global en el genoma. Las rutas biológicas enriquecidas obtenidas a través de los ARNm asociados a los ARNlnc, tales como una regulación positiva de la proliferación celular y una regulación negativa del proceso apoptótico.

tico, señalaban a un posible perfil oncogénico. Específicamente, *TCL6* fue el ARNlnc que mostró la mayor sobreexpresión en el subgrupo positivo para el gen de fusión *ETV6-RUNX1* (~4 veces). Según los análisis de co-expresión de ARNlnc y ARNm, la expresión de *TCL6* estuvo significativamente correlacionada con la expresión génica de *TCL1B* en nuestra propia cohorte y en varios conjuntos de datos externos. Los niveles de su ARNm asociado *TCL1B* también estuvieron sobre-expresados en pacientes con LLA-B positivos para *ETV6-RUNX1*.

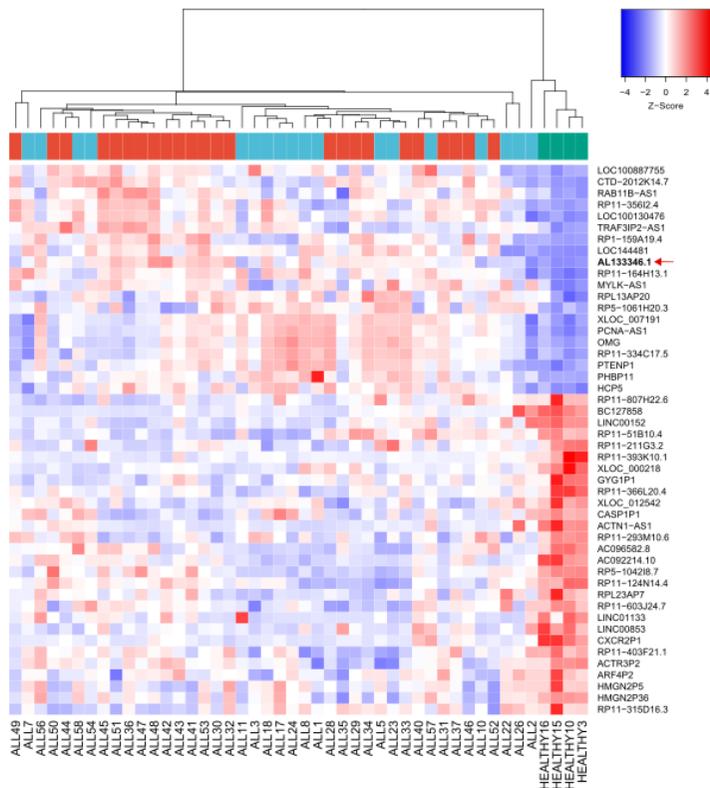
No solo *TCL6* fue identificado como un posible biomarcador ARNlnc para el diagnóstico de esta enfermedad, sino también se propuso como un factor pronóstico de los pacientes con LLA-B. Los individuos que tenían una menor expresión de *TCL6* presentaron una menor supervivencia libre de enfermedad de manera significativa ( $p$ -valor = 0,025), tanto en los casos negativos como positivos para la translocación *ETV6-RUNX1*. Aunque la función molecular del ARNlnc *TCL6* se desconoce en cáncer, la proteína *TCL1B* es activadora de la proteína quinasa *AKT* en leucemia de células T y posee actividad oncogénica. Este estudio es un ejemplo de cómo a partir del análisis de co-expresión de ARNlnc/ARNm previamente realizado se puede inferir una posible conexión funcional biológica que ayude a justificar el efecto en la supervivencia del paciente.

A pesar de conocerse en la bibliografía ARNlnc desregulados en las neoplasias hematológicas, son necesarios más estudios para llegar a obtener una firma de ARNlnc clínicamente relevante y definida para la LLA-B pediátrica. La identificación de los perfiles de ARNlnc/ARNm desregulados podría tener importantes repercusiones en esta enfermedad: proporcionar un diagnóstico más preciso, brindar una comprensión más profunda de la enfermedad y, en última instancia, ofrecer un mejor análisis del pronóstico para los pacientes.

## 3.2 Resultados

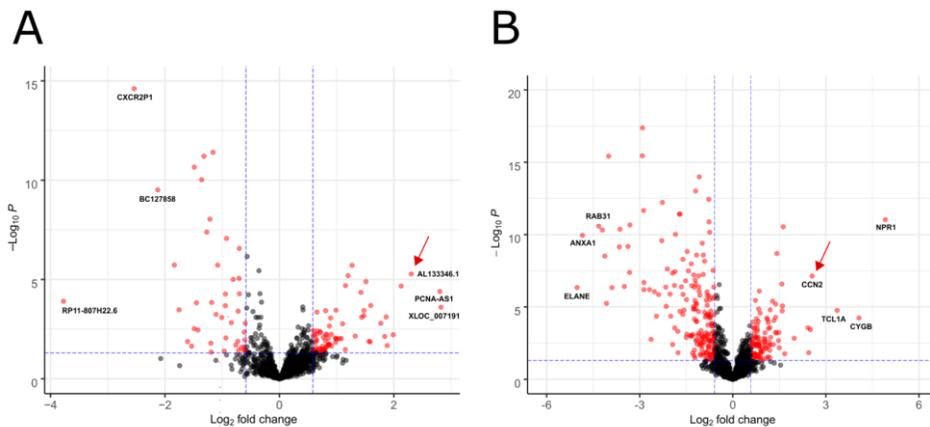
### 3.2.1 Firma de expresión génica aberrante en LLA-B pediátrica

En primer lugar, para obtener el perfil de expresión génica de ARNlnc en LLA-B pediátrica, se realizó una comparación mediante un *clustering* jerárquico entre los datos transcriptómicos de las 42 muestras de pacientes pediátricos con LLA-B y las 4 muestras de médula ósea de pacientes sanos obtenidos a partir del *LncPath<sup>TM</sup> Human Cancer Array*. A pesar de la heterogeneidad molecular característica de la LLA-B pediátrica, la comparativa realizada mostró un perfil de expresión génica diferencial que permitió agrupar de manera separada las muestras tumorales y las muestras sanas [Figura 9].



**Figura 9.** Heatmap de ARNlnc expresados diferencialmente entre las muestras de pacientes con LLA-B y las muestras de médula ósea sanas ( $|FC| > 1,5$ ,  $FDR < 0,05$ ). En la parte superior, la leyenda corresponde de la siguiente manera: rojo, muestras de LLA-B *ETV6-RUNX1* (+); azul, muestras de LLA-B *ETV6-RUNX1* (-); y verdes, muestras de médula ósea sanas. El *clustering* fue realizado de acuerdo con el coeficiente de correlación de Spearman. A la derecha, se encuentra señalado con una flecha roja el ARNlnc AL133346.1.

Estableciendo como umbrales un valor absoluto de FC > 1,5 y de FDR < 0,05, se identificaron 48 ARNInc diferencialmente expresados de manera significativa en las muestras tumorales respecto a las muestras sanas, de los cuales 20 se encontraban sobre-expresados y 28 regulados a la baja en los pacientes pediátricos con LLA-B [Tabla S1]. Por un lado, los ARNInc más sobre-expresados fueron *XLOC\_007191* (~7 veces), *PCNA-AS1* (~7 veces) y *AL133346.1* (~6 veces), mientras que los que se encontraban más regulados a la baja fueron *RP11-807H22.6* (~14 veces), *CXCR2P1* (~6 veces) y *BC127858* (~4 veces) [Figura 10A]. En cuanto a los ARNm analizados, se identificaron 233 diferencialmente expresados de manera significativa en las muestras de LLA-B, de los cuales 78 se encontraban sobre-expresados y 155 regulados a la baja. Entre aquellos más sobre-expresados, se encuentran *NPRI* (~30 veces), *CYGB* (~17 veces) y *TCL1A* (~10 veces), mientras que los que se encontraban más regulados a la baja fueron *ELANE* (~32 veces), *ANXA1* (~29 veces) y *RAB31* (~20 veces) [Figura 10B].



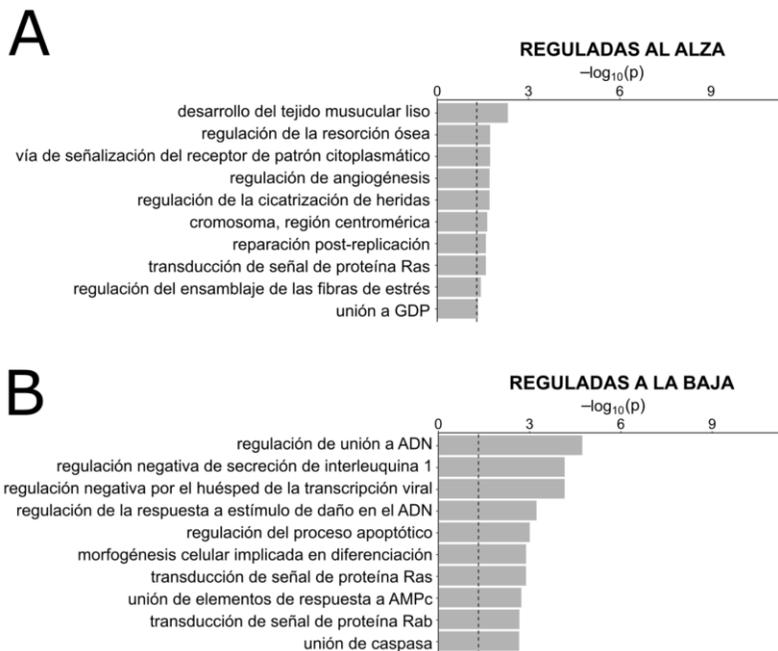
**Figura 10.** Volcano plots de ARNInc (A) y genes codificantes de proteínas (PCG) (B) expresados diferencialmente entre las muestras de pacientes pediátricos con LLA-B y las muestras de médula ósea sanas. La línea horizontal punteada azul representa un umbral FDR = 0,05, mientras que las líneas verticales representan los umbrales para el FC = -1,5/+1,5. Los puntos rojos representan aquellos ARNInc y PCG que superan los valores umbrales establecidos. Las flechas rojas de ambas gráficas indican la ubicación del ARNInc *AL133346.1* (izquierda) y su PCG asociado *CCN2* (derecha).

### 3.2.2 Rutas biológicas asociadas al perfil de ARNInc diferencialmente expresados en LLA-B

Tras el análisis de expresión génica diferencial, se estudiaron las posibles funciones biológicas alteradas asociadas al perfil de expresión génica de ARNInc obtenido. Debido a la incompleta anotación funcional de los ARNInc en las bases de datos disponibles, se utilizaron los PCG predichos, según la empresa diseña-

dora del *microarray*, como asociados funcionalmente a estos ARNInc expresados diferencialmente de forma significativa.

Los términos de la Ontología de Genes que se encontraban sobrerrepresentados estaban relacionados con la regulación de la angiogénesis (p-valor = 0,019; genes relacionados: *ANGPTL4*, *NF1*, *TNFAIP3*), la reparación del ADN (p-valor = 0,025; genes relacionados: *PCNA*, *UBE2VI*) y la ruta de transducción de señales de la proteína RAS (p-valor = 0,025; genes relacionados: *NF1*, *RAB4A*, *RAB11B*) [Figura 11A]. Por otro lado, los términos de la Ontología de Genes que se encontraban subrepresentados se relacionaban con la regulación negativa de la secreción de interleuquina 1 (p-valor =  $7,36 \times 10^{-5}$ ; genes relacionados: *CARD17*, *CARD18*), la regulación de la respuesta a la señal de daño en el ADN (p-valor =  $6,20 \times 10^{-4}$ ; genes relacionados: *HMG2*, *FEM1B*) y la regulación del proceso apoptótico (p-valor =  $1,04 \times 10^{-3}$ ; genes relacionados: *ACTN1*, *EPHA1*, *HMG2*, *JUN*, *NOTCH2*) [Figura 11B]. En conjunto, estos resultados sugieren una firma oncogénica global asociada con los perfiles de expresión de ARNInc/ARNm en las muestras de LLA-B pediátrica en comparación con las muestras de médula ósea sanas.



**Figura 11.** Análisis de ontología génica (GO) de los ARNm asociados con los ARNInc expresados diferencialmente de manera significativa entre las muestras de LLA-B pediátricas y las muestras de médula ósea sanas. Las gráficas muestran los diez términos GO más enriquecidos de manera positiva (A) y de manera negativa (B).

### 3.2.3 Altos niveles de expresión génica de la pareja *AL133346.1/CCN2* en LLA-B

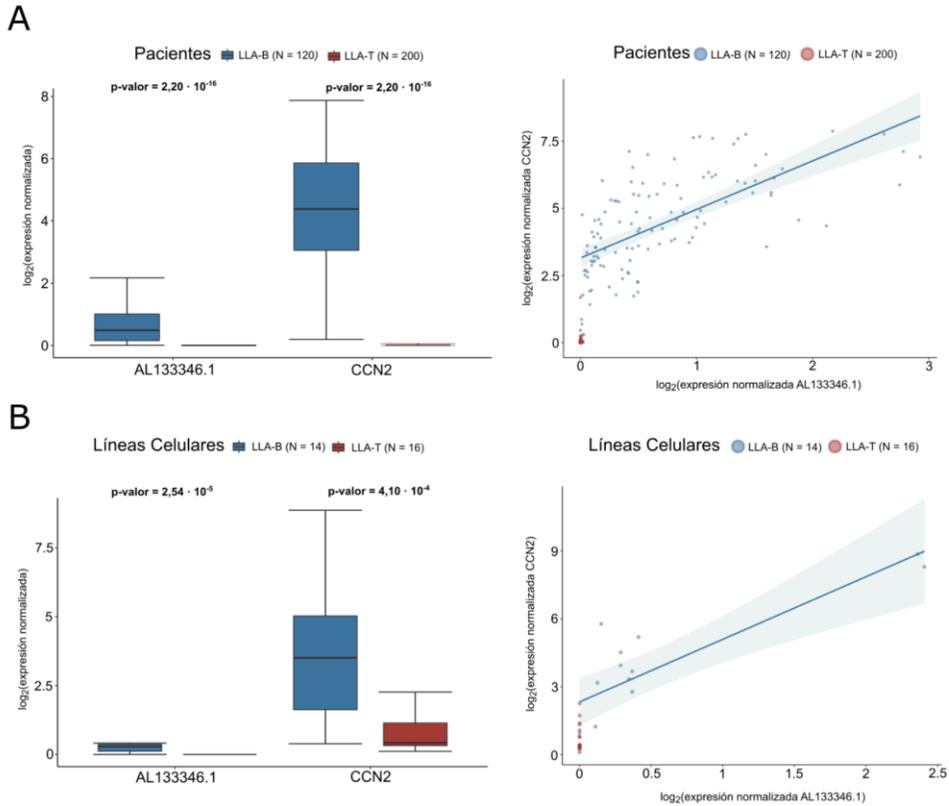
Dentro de las parejas de ARNInc/ARNm que se encontraban diferencialmente expresadas entre los pacientes pediátricos con LLA-B y las muestras médula ósea sanas ( $|FC| > 1,5$ ,  $FDR < 0,05$ ), *AL133346.1/CCN2* se encontraban sobre-expresados en los pacientes positivos para la translocación *ETV6-RUNX1* en comparación con los negativos en el anterior estudio (Cuadros et al., 2019). En los resultados pertenecientes a este Capítulo III, dentro de los grupos de expresión génica regulados al alza en pacientes pediátricos con LLA-B, *AL133346.1* era el tercer ARNInc más sobre-expresado ( $FC = 5,16$ ,  $FDR = 1,48 \times 10^{-3}$ ) [Figura 10A, Tabla S1], mientras que *CCN2* se encontraba en cuarta posición en los PCG ( $FC = 5,24$ ,  $FDR = 3,00 \times 10^{-4}$ ) [Figura 10B]. Por ello, debido a sus patrones de expresión génica, se decidió analizar en mayor profundidad los niveles de *AL133346.1* y *CCN2* en otros conjuntos de datos adicionales.

Con el fin de confirmar la expresión aberrante *AL133346.1* y *CCN2* en otras cohortes externas de mayor tamaño, se utilizaron los datos pertenecientes al proyecto TARGET, específicamente de los proyectos TARGET-ALL-P1 y TARGET-ALL-P2, que contienen información ómica sobre pacientes con LLA. Una vez descartados aquellos pacientes mayores de 14,99 años y que no poseían información disponible para la expresión de *AL133346.1* y *CCN2*, se obtuvo una población compuesta por 320 pacientes pediátricos, 120 de ellos con LLA-B y 200 de ellos con LLA-T. Dado que en aquel momento no existían datos disponibles públicos de pacientes pediátricos sanos en los mismos proyectos, se realizó una comparación de la expresión de ambos genes entre las cohortes de pacientes de los dos subtipos de LLA con el fin de demostrar la especificidad de los niveles de *AL133346.1* y *CCN2* en el tipo celular B. Como resultado, se obtuvo una mayor expresión en los pacientes con LLA-B en comparación con aquellos con LLA-T tanto para *AL133346.1* ( $p\text{-valor} = 2,20 \times 10^{-16}$ ) como para *CCN2* ( $p\text{-valor} = 2,20 \times 10^{-16}$ ) [Figura 12A].

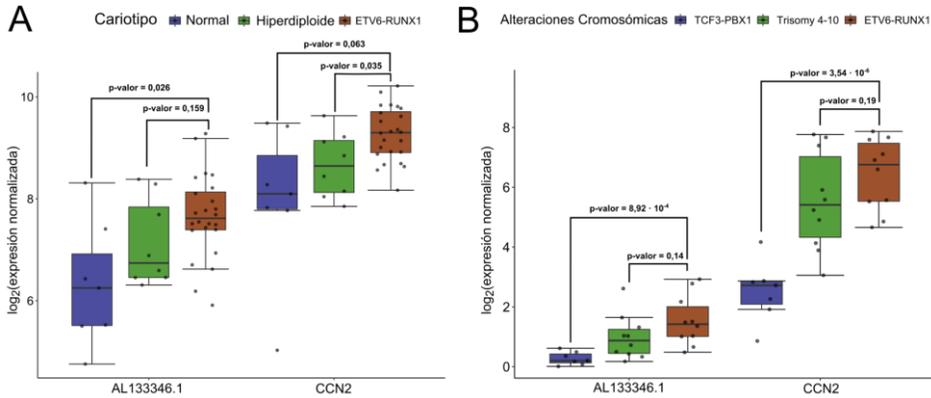
La expresión específica de *AL133346.1* y *CCN2* en pacientes pediátricos con LLA-B fue confirmada realizando la comparación con el subtipo celular T con datos relativos a líneas celulares procedentes de la base de datos CCLE. Como resultado, se obtuvo una mayor expresión en las líneas celulares de LLA-B en comparación con aquellas de LLA-T tanto para *AL133346.1* ( $p\text{-valor} = 2,54 \times 10^{-5}$ ) como para *CCN2* ( $p\text{-valor} = 4,10 \times 10^{-4}$ ) [Figura 12B].

Por último, dentro de los subtipos de LLA-B pediátrica asociados a los distintos tipos de translocaciones conocidos, *AL133346.1* y *CCN2* mostraron su

mayor nivel de expresión en aquellos pacientes positivos para la translocación *ETV6-RUNX1* en comparación con aquellos pacientes con hiperdiploidía y con cariotipo normal pertenecientes a nuestra cohorte del *microarray* de expresión [Figura 13A]. En el caso de la cohorte del proyecto TARGET, únicamente se obtuvo una mayor expresión estadísticamente significativa en comparación con los pacientes con la fusión génica *TCF3-PBX1* tanto para *AL133346.1* (p-valor =  $8,92 \times 10^{-4}$ ) como para *CCN2* (p-valor =  $3,54 \times 10^{-6}$ ) [Figura 13B].



**Figura 12.** Comparación de la expresión génica de *AL133346.1/CCN2* entre las muestras de LLA-B y LLA-T derivadas de los pacientes pediátricos del proyecto TARGET de TCGA (A) y de las líneas celulares de la CCLE (B). A la izquierda, en los *boxplots* se compara la expresión génica normalizada de ambos genes entre ambos subtipos B y T, junto a los p-valores correspondientes. A la derecha, en los *scatterplots* se representan los niveles de expresión de ambos genes para analizar su correlación. Los tamaños de muestra y los coeficientes de correlación de Spearman son: **Pacientes:** 120 LLA-B ( $\rho = 0,717$ , p-valor < 0,0001) vs. 200 LLA-T ( $\rho = 0,099$ , p-valor = 0,163). **CCLE:** 14 LLA-B ( $\rho = 0,799$ , p-valor = 0,001) vs. 16 LLA-T (NA).



**Figura 13.** Comparación de la expresión génica de *AL133346.1/CCN2* entre los distintos grupos de pacientes pediátricos con LLA-B con diferentes cariotipos o alteraciones cromosómicas de nuestro *microarray* (A) o del proyecto TARGET de TCGA (B). Los *box-plots* están coloreados dependiendo del subtipo de LLA-B.

### 3.2.4 Expresión de *AL133346.1/CCN2* correlacionada en LLA-B

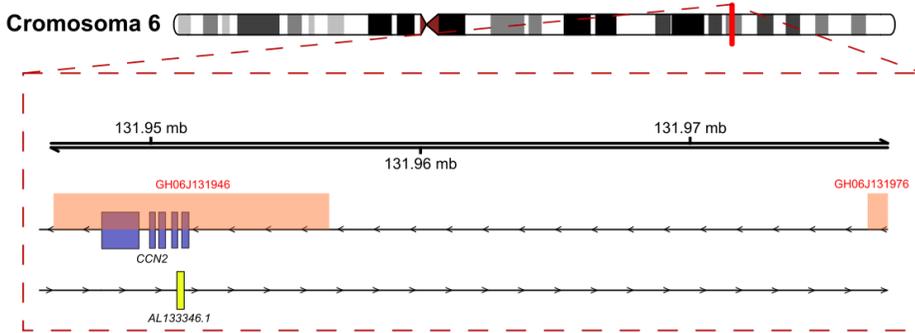
Los ARNlnc pueden regular los niveles de ARNm a través de distintos mecanismos estableciendo un posible patrón de correlación de expresión génica entre el ARNlnc y sus ARNm asociados (Y. Zhang et al., 2016). En el anterior estudio, se demostró una correlación positiva significativa entre los niveles de ARN de *AL133346.1* y *CCN2* tanto en nuestra cohorte como en tres distintos *microarrays* de otros estudios sobre LLA-B pediátrica (Cuadros et al., 2019). Con el fin de confirmar esta correlación, se obtuvieron los coeficientes de correlación de Spearman de los niveles de expresión de *AL133346.1* y *CCN2* de los datos procedentes del proyecto TARGET y de la CCLE. En cuanto a las muestras de pacientes, se obtuvo una correlación positiva significativa en aquellos con LLA-B (N = 120,  $\rho = 0,717$ , p-valor < 0,0001), mientras que en el caso de LLA-T el resultado no fue estadísticamente significativo (N = 200,  $\rho = 0,099$ , p-valor = 0,163) debido a la expresión prácticamente nula del ARNlnc *AL133346.1* [Figura 12A]. De manera similar, se obtuvo una correlación positiva significativa para las líneas celulares de LLA-B de la base de datos CCLE (N = 14,  $\rho = 0,799$ , p-valor = 0,001) a diferencia de las líneas celulares de LLA-T (N = 16,  $\rho = \text{NA}$ , p-valor = NA) que no expresan el ARNlnc *AL133346.1* [Figura 12B]. Así pues, estos resultados sugieren que la correlación de la expresión génica de la pareja *AL133346.1/CCN2* es específica del subtipo celular B.

### 3.2.5 Análisis *in silico* de los mecanismos de regulación de *AL133346.1/CCN2*

Para evaluar si *AL133346.1* podría estar involucrado en la regulación de otros genes, se estudiaron las correlaciones de la expresión génica de *AL133346.1* con la expresión del resto de PCG incluidos en nuestro *microarray* y en los datos procedentes de TARGET de pacientes pediátricos con LLA-B. En nuestro *microarray*, que incluía sondas para 1.906 PCG, la expresión de *AL133346.1* únicamente estaba correlacionada de manera significativa con la expresión de *CCN2* ( $\rho = 0,840$ , p-valor =  $3,06 \times 10^{-13}$ ). En los datos procedentes de TARGET de tecnología RNA-Seq, usando todos los PCG disponibles, 9.080 mostraron una correlación estadísticamente significativa con la expresión de *AL133346.1*, donde *CCN2* estaba entre los 100 primeros con mayor poder estadístico.

Con el fin de confirmar que la expresión de *AL133346.1* produce un efecto en la expresión proteica de *CCN2*, se decidió estudiar la correlación de los niveles de ARN y proteína de *CCN2* en datos procedentes de las 27 muestras de pacientes pediátricos con LLA-B del estudio de Yang et al (M. Yang et al., 2019). Como resultado, la expresión de ARN y proteína de *CCN2* se encontraban correlacionadas significativamente de manera moderada ( $\rho = 0,518$ , p-valor = 0,001). De esta manera, este resultado sugiere que los cambios en la expresión génica de *AL133346.1* podrían tener un impacto en la expresión proteica de *CCN2*.

Por último, se decidió analizar el *locus* genómico donde se encuentran los genes *AL133346.1* y *CCN2* para encontrar posibles promotores o zonas regulatorias cercanas que pudieran explicar los niveles de correlación estudiados. Ambos genes se encuentran en el cromosoma 6 (q23.2), en diferentes hebras orientados cabeza-cabeza, superpuestos parcialmente y con sus TSS separados por 2.770 pb. Según la base de datos *GeneHancer* (Fishilevich et al., 2017), que integra información de promotores y *enhancers* de otras fuentes, *AL133346.1* y *CCN2* comparten dos de sus promotores/*enhancers* de alta puntuación predichos: GH06J131946 y GH06J131976 [Figura 14]. Estos dos *enhancers* se clasifican como *enhancers* de alta probabilidad, anotados como *elite* en la base de datos, y muestran una fuerte asociación potenciador-gen.



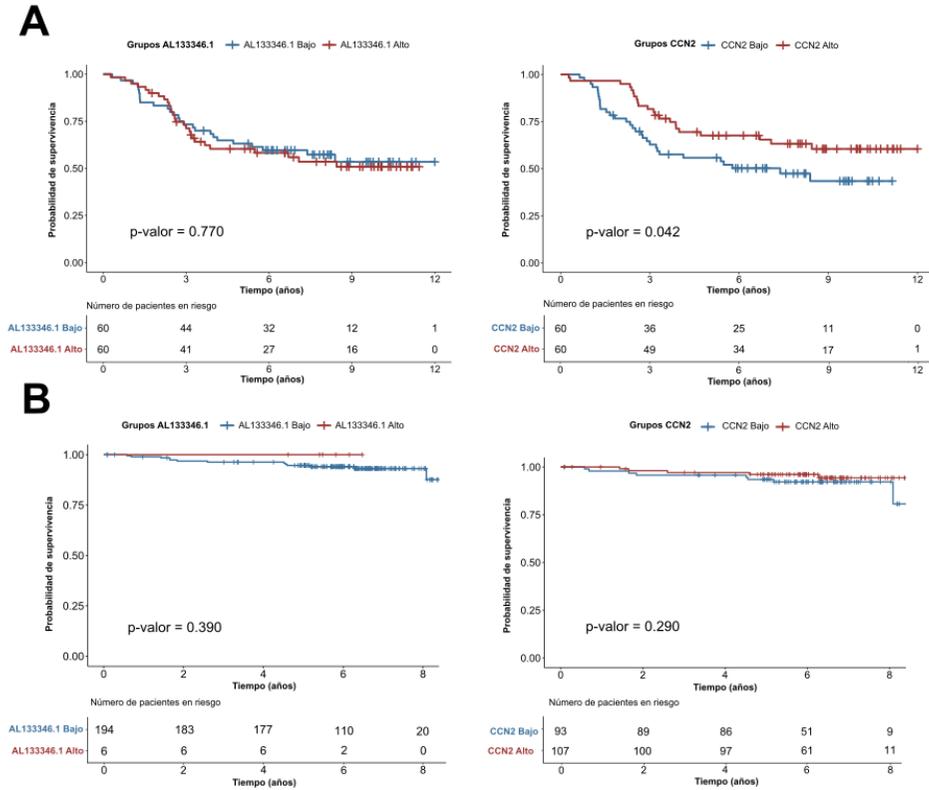
**Figura 14.** Representación esquemática de la localización genómica de *AL133346.1/CCN2* en el cromosoma 6. El cuadro con rayas discontinuas rojas representa un aumento de *zoom* en el *locus*. Por motivos de optimización de la visualización, se encuentran representados todos los exones de *CCN2* (azul) y únicamente el primero de *AL133346.1* (amarillo). Las regiones reguladoras GH06J131946 y GH06J131976 se encuentran representadas con cuadros rojos. Abreviaturas: mb: Megabases.

### 3.2.6 Altos niveles de expresión de *CCN2* asociados con un mejor pronóstico en pacientes con LLA-B

Con el fin de determinar el valor pronóstico que poseen la expresión de *AL133346.1* y *CCN2*, se realizó un análisis de supervivencia utilizando datos clínicos de pacientes pediátricos con LLA-B y con LLA-T procedentes de la base de datos TARGET. Las variables clínicas obtenidas fueron el sexo, la edad en el momento del diagnóstico, la afectación del SNC, las translocaciones t(12;21)[*ETV6/RUNX1*], t(1;19)[*TCF3/PBX1*] y t(9;22)[*BCR/ABL1*], el reordenamiento del gen *MLL*, las trisomías del cromosoma 4 y 10, y el número de cromosomas o ploidía [Figura S1; Figura S2; Tabla S2].

En cuanto a las muestras del subtipo celular B, el análisis de Cox univariable realizado para las covariables clínicas obtenidas mostró que el sexo masculino (p-valor = 0,030), la ploidía (Hipodiploidía: p-valor = 0,005; Hiperdiploidía Parcial: p-valor = 0,033; Hiperdiploidía Alta: p-valor = 0,019) estaban significativamente asociados con una supervivencia global más corta de los pacientes, mientras que la presencia de la fusión génica *TCF3-PBX1* (p-valor =  $1,22 \times 10^{-7}$ ) se encontraba significativamente asociada a un mayor riesgo de muerte. Sin embargo, la edad en el momento del diagnóstico (p-valor = 0,186), la afectación del SNC (p-valor = 0,943), la presencia de la fusión génica *ETV6-RUNX1* (p-valor = 0,490), la fusión génica *BCR-ABL1* (p-valor = 0,689) y la trisomía de los cromosomas 4 y 10 (p-valor = 0,418), no fueron consideradas variables predictoras significativas de la supervivencia de esta cohorte [Tabla 6]. Al dividir a los pacientes por la mediana de expresión génica de *AL133346.1* y *CCN2*, los pacientes

con mayor expresión de *CCN2* estaban asociados con tiempos de supervivencia globales más largos (p-valor = 0,042; HR: 0,566; IC del 95%: 0,328-0,980), mientras que no hubo asociación estadísticamente significativa entre la expresión de *AL133346.1* y la supervivencia global (p-valor = 0,770) [Figura 15A].



**Figura 15.** Curvas de supervivencia global Kaplan-Meier de los pacientes pediátricos con LLA-B (A) y LLA-T (B) de acuerdo con la división de los grupos de expresión de los genes *AL133346.1* (izquierda) y *CCN2* (derecha) en “Bajo” y “Alto”. Se muestran los p-valoros asociados a los análisis de supervivencia.

Todas las covariables clínicas con p-valor < 0,2 en el análisis de Cox univariable, junto con la covariable relacionada con la expresión de *CCN2*, se utilizaron para un análisis de Cox multivariable. La fusión génica *TCF3-PBX1* fue la variable con mayor influencia en la supervivencia global de pacientes pediátricos con LLA-B (p-valor = 0,003; HR: 7,589; IC del 95%: 1,914-30,085). Además, este análisis confirmó que la alta expresión de *CCN2* podría ser un factor pronóstico independiente asociado con una supervivencia global más larga (p-valor = 0,045; HR: 0,448; IC del 95%: 0,204-0,984) [Tabla 1].

**Tabla 6.** Análisis Cox univariable y multivariable de las covariables clínicas de los pacientes pediátricos con LLA-B derivados del proyecto TARGET de TCGA (N = 120). Los valores NA aparecen debido a que las correspondientes covariables clínicas no fueron incluidas en el Análisis Cox Multivariable. La covariable relacionada con los reordenamientos *MLL* no fue incluida debido a que no existían pacientes positivos.

	Análisis Cox Univariable		Análisis Cox Multivariable	
	Cociente de Riesgo (95% IC)	p-valor	Cociente de Riesgo (95% IC)	p-valor
<b>Sexo (Hombre vs Mujer)</b>	1,852 (1,062-3,228)	0,030	1,900 (0,908-3,973)	0,088
<b>Edad al Diagnóstico (≥ 10 vs 1-9.9 años)</b>	1,500 (0,823-2,733)	0,186	1,140 (0,510-2,550)	0,750
<b>CNS (2-3 vs 1)</b>	1,024 (0,538-1,949)	0,943	NA	NA
<b>t(12;21): ETV6-RUNX1 (Positivo vs Negativo)</b>	0,697 (0,250-1,944)	0,490	NA	NA
<b>t(1;19): TCF3-PBX1 (Positivo vs Negativo)</b>	21,840 (6,968-68,480)	$1,22 \times 10^{-7}$	7,589 (1,914-30,085)	0,003
<b>t(9;22): BCR-ABL1 (Positivo vs Negativo)</b>	0,667 (0,092-4,834)	0,689	NA	NA
<b>Trisomías 4 y 10 (Positivo vs Negativo)</b>	0,343 (0,192-1,986)	0,418	NA	NA
<b>Ploidía (Hipodiploidía vs Diploidía)</b>	0,126 (0,030-0,531)	0,005	0,237 (0,030-1,881)	0,173
<b>Ploidía (Hiperdiploidía Parcial vs Diploidía)</b>	0,383 (0,158-0,926)	0,033	0,650 (0,252-1,679)	0,374
<b>Ploidía (Hiperdiploidía Alta vs Diploidía)</b>	0,346 (0,143-0,838)	0,019	0,386(0,148-1,006)	0,051
<b>Expresión CCN2 (Alto vs Bajo)</b>	0,566 (0,328-0,980)	0,042	0,448 (0,204-0,984)	0,045
<b>Expresión AL133346.1 (Alto vs Bajo)</b>	1,084 (0,632-1,859)	0,770	NA	NA

Por último, para estudiar si el nivel de expresión de *CCN2* podría ser un biomarcador pronóstico específico de LLA-B pediátrica, se realizó un análisis de supervivencia utilizando los datos clínicos de los pacientes con LLA-T. En este caso, no hubo asociación entre la expresión de *AL133346.1* (p-valor = 0,390) y la expresión de *CCN2* (p-valor = 0,290) con la supervivencia global en pacientes pediátricos con LLA-T [Figura 15B], sugiriendo que estos genes afectan de manera específica a la supervivencia de los pacientes con subtipo B.

## 4. Discusión

Numerosos estudios han descubierto ARNlnc asociados tanto al inicio como al progreso de un amplio rango de enfermedades (Wapinski & Chang, 2011). Específicamente, en cáncer existen ARNlnc que pueden actuar como oncogenes o como supresores tumorales (Rinn & Chang, 2012). Por ejemplo, la sobre-expresión de *HOTAIR* se ha relacionado con la metástasis del cáncer de mama (Gupta et al., 2010). A pesar del descubrimiento de miles de nuevos ARNlnc gracias a las técnicas de secuenciación masiva, solo un pequeño porcentaje se encuentran relacionados con el pronóstico y el diagnóstico de enfermedades. Por ello, la mejora de estas técnicas de detección se ha complementado con el desarrollo de nuevos *softwares* que persiguen diagnosticar enfermedades a partir de distintas fuentes de información, como la expresión similar de ARNlnc, las secuencias génicas o el fenotipo del paciente. Por ejemplo, la herramienta bioinformática LncDisease detecta asociaciones globales de ARNlnc con enfermedades mediante la posible interacción con miARN (J. Wang et al., 2016). En relación con los resultados presentados en este Capítulo III, existen otros estudios con metodología basada en la co-expresión de ARNlnc/ARNm, utilizada para predecir la función de ARNlnc y su posible asociación con patologías (Liao et al., 2011; L. Wang et al., 2014).

En comparación con los ARNm, los ARNlnc muestran niveles de expresión génica más bajos, poseen tiempos de vida medio más cortos y presentan una mayor especificidad celular (Clark et al., 2012). Además, los bajos niveles de ARNlnc detectados en análisis de secuenciación de ARN-Seq pueden llegar a ser 10 veces menor que los obtenidos para los ARNm de PCG debido entre otras razones, a la baja profundidad de secuenciación o la selección específica de transcritos con poli-A (Y. Guo et al., 2015; Uszczyńska-Ratajczak et al., 2018), lo que dificulta la cuantificación exacta de ARNlnc. En nuestro estudio, se recurrió al *LncPath<sup>TM</sup> Human Cancer Array* para detectar de manera más precisa los ARNlnc en comparación con las tecnologías mencionadas de secuenciación de ARN-Seq.

En estudios previos, se ha demostrado que la expresión de los ARNlnc en LLA-B se encuentra asociada de manera específica a las características citogenéticas de los pacientes. Fernando et al. demostraron un perfil diferencial de expresión de ARNlnc en muestras de pacientes pediátricos con LLA-B asociados a las fusiones génicas *ETV6-RUNX1*, *TCF3-PBX1* y *MLL-AF4* (Fernando et al., 2015). En nuestro estudio, el análisis comparativo de los niveles de expresión de ARNlnc del *microarray* permitió discriminar y agrupar las muestras de los pacientes pediátricos con LLA-B y las muestras de médula ósea sanas de manera

separada. Sin embargo, se debe tener en cuenta que la cohorte pediátrica analizada no tenía presente ningún otro tipo de translocación diferente a la que produce el gen de fusión *ETV6-RUNX1*, por lo que el perfil de ARNlnc obtenido en comparación con las muestras sanas es específico de esta translocación. Dado el relevante papel regulatorio que poseen los ARNlnc en procesos de diferenciación y carcinogénesis, el análisis de rutas biológicas enriquecidas en los pacientes pediátricos con LLA-B mostró un patrón global relacionado con la oncogénesis.

Un hallazgo relevante de este Capítulo III es la propuesta de la expresión del ARNlnc *AL133346.1* como un potencial biomarcador diagnóstico de la LLA-B pediátrica, tanto para distinguir de pacientes sanos como de pacientes pediátricos con LLA-T. *AL133346.1*, previamente conocido como *RP11-69I8.3*, es un ARNlnc que había sido descubierto regulado a la baja en leiomiomas uterinos (Chuang et al., 2022). En el análisis de este Capítulo III, el ARNlnc se encontraba sobre-expresado significativamente en el subtipo celular B en comparación con las muestras de médula ósea sanas y los pacientes pediátricos con LLA-T procedentes del proyecto TARGET. Además, esta sobre-expresión específica también se pudo obtener en su PCG asociado *CCN2*. Sin embargo, ya existían estudios previos que demuestran que la sobre-expresión de *CCN2* es exclusiva del subtipo celular pre-B (Boag et al., 2007).

En cuanto a los mecanismos de regulación de los ARNlnc, si bien la expresión génica de éste y su gen codificador de proteína vecino pueden estar correlacionadas, el ARNlnc puede no estar implicado directamente en la regulación del segundo. Diversos estudios han demostrado que la función reguladora de ciertos genes de ARNlnc puede no residir en el producto de ARN generado a partir de este sino en los elementos reguladores contenidos dentro del gen del ARNlnc. Por ejemplo, se pensaba inicialmente que el ARNlnc *Lockd*, ubicado de manera cercana a *Cdkn1b*, modulaba la expresión del PCG en *cis* en líneas celulares eritroides de ratón. Sin embargo, experimentos posteriores demostraron que la expresión del ARNlnc *Lockd* era probablemente un subproducto transcripcional que no modulaba la expresión de *Cdkn1b*. Los autores concluyeron que la actividad biológica del *locus* genómico de ambos genes residía en un elemento *enhancer* cercano (Paralkar et al., 2016). Por ello, aunque los niveles de expresión de *AL133346.1* y su PCG asociado *CCN2* se correlacionen, es necesario confirmar mediante experimentos *in vitro* adicionales su posible relación funcional. Así pues, cabe la posibilidad de que un factor de transcripción común pueda modular la expresión tanto del ARNlnc como del gen codificador de proteína vecino a través de su interacción con estos elementos reguladores, generando así un nivel transcripcional correlacionado de ambos genes (Kopp & Mendell, 2018; Paralkar et al., 2016). Para evaluar si este fenómeno puede explicar la co-

expresión de *AL133346.1* y *CCN2*, se estudió la ubicación genómica de ambos genes y se realizó una búsqueda de promotores cercanos y regiones regulatorias próximas. Los resultados de correlación positiva de genes comentados en la anterior sección LLA-B y las regiones reguladoras encontradas apuntan a la posibilidad de que el gen codificador de proteína *CCN2* y su ARNlnc antisentido *AL133346.1* se expresen de manera específica en el tejido cuando su *locus* se encuentra activo transcripcionalmente. De manera global, según nuestros resultados, se proponen dos hipótesis mecanísticas principales no excluyentes que apoyarían este modelo: 1) *AL133346.1* regula la expresión de *CCN2* en *cis*; o 2) *AL133346.1* y *CCN2* son modulados específicamente en LLA-B por los mismos elementos reguladores.

Para evaluar el valor pronóstico que podría poseer la pareja *AL133346.1/CCN2* descubierta en el análisis de expresión, se utilizaron datos clínicos del proyecto TARGET debido a que los pacientes sanos disponibles en nuestro estudio no generaban subgrupos lo suficientemente amplios en el análisis de supervivencia para obtener resultados significativos. Las curvas de supervivencia Kaplan-Meier mostraron de manera específica para las muestras pediátricas con LLA-B que el gen *CCN2* influye de manera significativa en la supervivencia global del paciente. Sin embargo, no ocurrió de la misma manera para el ARNlnc *AL133346.1*, posiblemente debido a la baja sensibilidad que las técnicas de RNA-Seq suelen ofrecer para la detección de este tipo de ARN. Además, la correlación positiva de *AL133346.1/CCN2* obtenida para la cohorte de TARGET ( $\rho = 0,717$ ) fue menor que la del microarray ( $\rho = 0,840$ ), por lo que puede provocar que la separación de los grupos en el análisis de supervivencia según la expresión del ARNlnc y del ARNm no sea la misma. Además, se debe tener en cuenta que dos de los parámetros clínicos más fiables para determinar la agresividad, las translocaciones  $t(12;21)[ETV6/RUNX1]$  y  $t(1;19)[TCF3/PBX1]$  (Kang et al., 2010) no se distribuyeron homogéneamente en los subgrupos analizados. Los subgrupos de expresión "Alto" de *AL133346.1* y *CCN2* incluían el 90% y el 100% de las muestras positivas para  $t(12; 21)[ETV6-RUNX1]$ , mientras que poseían el 29% y el 0% de los casos positivos para la  $t(1;19)[TCF3/PBX1]$ , respectivamente. Aun así, según el análisis Cox multivariable, la expresión de *CCN2* fue considerada como un factor pronóstico independiente de la cohorte analizada.

*CCN2* (del inglés, “*Cellular Communication Network Factor 2*”), también conocido como CTFG (del inglés, “*Connective Tissue Growth Factor*”), pertenece a la familia de proteínas de la red de comunicación celular CCN (Hashiguchi et al., 2022). Esta familia de proteínas asociadas a la matriz extracelular se caracteriza por tener cuatro dominios conservados ricos en cisteína, los

cuales interactúan con distintos receptores de superficie de la matriz estructural, proteasas y factores de crecimiento (Perbal, 2013). *CCN2* es una proteína multifuncional implicada en diversos procesos fisiológicos, tales como la angiogénesis, el desarrollo tumoral, las enfermedades fibrosas y la cicatrización de heridas (Kubota & Takigawa, 2015). Específicamente, esta proteína se encuentra altamente expresada en células endoteliales ya que su principal función es promover la angiogénesis (Ivkovic et al., 2003). Así, en cáncer de próstata y de mama, las células tumorales secretan *CCN2* para promover la migración de las células endoteliales e inducir la angiogénesis tumoral (Shimo et al., 2001; F. Yang et al., 2005). En el contexto de las neoplasias hematopoyéticas, la familia *CCN* juega un papel relevante en la diferenciación de las HSC y las células madre mesenquimales (Lenz et al., 2008). Se ha demostrado que la expresión de *CCN2* de las células estromales desempeña un papel crucial en el mantenimiento y la supervivencia a largo plazo de las HSC, así como en la generación de células mieloides y células B (Leguit et al., 2021). El valor pronóstico de la familia *CCN* de proteínas varía en las diferentes neoplasias hematológicas (Wells et al., 2015). Así, apoyando los resultados obtenidos en este Capítulo III, varios estudios correlacionaron significativamente la expresión de *CCN2* con una mayor supervivencia global en el linfoma tratado con quimioterapia (Lenz et al., 2008; Rimsza et al., 2008). Sin embargo, en este contexto terapéutico, se necesitan estudios específicos adicionales en LLA-B pediátrica para determinar si los niveles de expresión de *CCN2* podrían influir en el resultado del tratamiento del paciente.

# **Capítulo IV**

## **Implicación de ARID2 en LLA**

# **CAPÍTULO IV: Implicación de ARID2 en LLA**

El Capítulo IV describe los resultados obtenidos relacionados con el Objetivo 2.1, basado en el análisis de las alteraciones genéticas de las subunidades del complejo SWI/SNF en cáncer. Específicamente, se estudió en mayor profundidad la subunidad ARID2 y su posible papel funcional en la LLA-B a través del análisis de datos ómicos disponibles en bases de datos y de diferentes validaciones experimentales realizadas en el laboratorio.

## **4.1 Antecedentes**

### **4.1.1 Funciones y alteraciones del complejo SWI/SNF en la hematopoyesis**

La función del complejo remodelador de la cromatina SWI/SNF es esencial para la regulación de los distintos procesos de diferenciación y mantenimiento de las HSC (Tu & Zheng, 2022; Z. Wang et al., 2021). Cada vez son más frecuentes los estudios que definen de manera específica cómo se encuentran involucradas las subunidades del complejo SWI/SNF en la homeostasis de los procesos hematológicos, produciendo así los cambios necesarios en el transcriptoma para controlar de manera específica los patrones de expresión génica de las células hematopoyéticas en sus distintos estadios de diferenciación (Andrades et al., 2023).

Distintas publicaciones recientes han señalado las funciones esenciales de varias subunidades del complejo SWI/SNF, como ARID1A, PBRM1, ARID2 y SMARCA2, en relación con el mantenimiento de las HSC (Bluemn et al., 2021; Han et al., 2019; Lee et al., 2016; Naidu et al., 2022). Por otro lado, de manera más específica, existen otras subunidades que se encuentran implicadas en tipos celulares concretos, tales como ARID1A y SMARCA4 en la diferenciación linfóide y mielóide (Astori et al., 2020; Gebuhr et al., 2003), ARID1B y SMARCA4 en la eritropoyesis (Azad et al., 2022) o SMARCD2 en la granulopoyesis (Priam et al., 2017). Las alteraciones genéticas descritas hasta la fecha en las subunidades implicadas en estos procesos hematológicos pueden conducir al inicio o a la progresión de un amplio abanico de neoplasias hematológicas. En general, los genes de las subunidades afectadas son considerados supresores tu-

morales dado que las alteraciones producidas generan una pérdida de función del complejo remodelador de cromatina y, en consecuencia, el inicio de la enfermedad. Sin embargo, para el mismo gen, existen casos donde una ganancia de función o su amplificación conduce a una aceleración de la patología, considerándose así oncogenes (Andrades et al., 2023).

Por último, es importante destacar que las alteraciones genéticas de las distintas subunidades del complejo SWI/SNF en una misma neoplasia hematológica pueden tener efectos funcionales completamente distintos. Esto se debe a que los complejos SWI/SNF residuales generados tras las alteraciones podrían manifestar diferentes actividades biológicas (Bluemn et al., 2021). Por ejemplo, la proteína de fusión oncogénica KMT2A-MLLT1 forma conjuntos anómalos de subunidades del complejo SWI/SNF denominados EBAFb (del inglés, "*ENL-associated BAF-containing BAF250b*"). Los subconjuntos EBAFb, que contienen la subunidad catalítica SMARCA4, activan de manera específica la transcripción de genes que se expresan de manera aberrante en la leucemia de linaje mixto (Nie et al., 2003).

### 4.1.2 El papel de ARID2 en la hematopoyesis.

Actualmente, pocos estudios se han centrado en la regulación funcional que ejercen los complejos PBAF, que contienen la subunidad ARID2, sobre el programa transcripcional del proceso de la hematopoyesis. Recientemente, se ha descubierto el rol funcional que posee en la limitación de la expansión de las células T CD8<sup>+</sup> en el contexto tumoral y en infecciones crónicas virales (Kharel et al., 2023). En anteriores estudios, *ARID2* ha sido considerado como un gen supresor tumoral de la metástasis de hepatocarcinoma y en melanoma (Carcamo et al., 2022; Loesch et al., 2020), aunque su implicación no ha sido analizada en profundidad en el contexto de las enfermedades hematológicas. Existen algunos estudios enfocados en dilucidar el papel de ARID2 en el proceso de hematopoyesis a través de su delección o inhibición en organismos modelo. Hasta ahora, los resultados fenotípicos de estos experimentos apuntan a una anemia leve durante la hematopoyesis, una disminución de la capacidad de diferenciación de las células progenitoras hematopoyéticas hacia la línea linfoide y una aceleración de la leucemogénesis en modelos de LMA con translocación MLL-AF9 (Bluemn et al., 2021; L. Liu et al., 2018).

Las alteraciones genéticas recurrentes en las subunidades del complejo SWI/SNF no solo subrayan la relevancia biológica de estos sistemas remodeladores de la cromatina en las neoplasias hematológicas, sino también su potencial clínico. Algunos estudios indican que las alteraciones en el complejo SWI/SNF generan resistencia a agentes antineoplásicos utilizados en el tratamiento de estas

## CAPÍTULO IV: Implicación de ARID2 en LLA

enfermedades (Andrades et al., 2023). Así pues, la modulación de la funcionalidad de los genes alterados del complejo podría ayudar a incrementar la sensibilidad a agentes quimioterapéuticos con el fin de mejorar el pronóstico de los pacientes. Dada la relevancia funcional del complejo PBAF en el control de la expresión génica, es importante caracterizar el rol biológico que ejerce ARID2 en el contexto de la leucemia con el fin de encontrar nuevas dianas terapéuticas en esta enfermedad.

## 4.2 Resultados

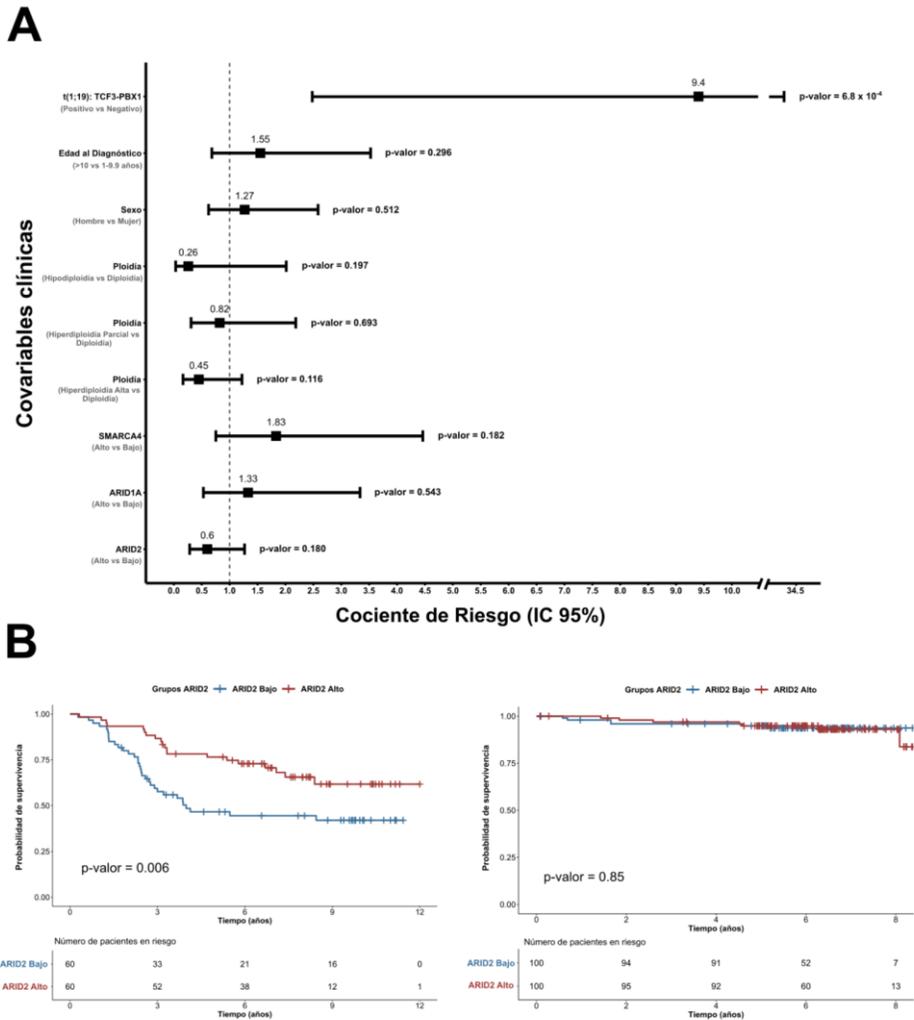
### 4.2.1 Valor pronóstico de *ARID2* en pacientes pediátricos con LLA-B

Inicialmente, con el fin de analizar qué subunidad del complejo SWI/SNF podría estar influyendo de manera independiente en la supervivencia de los pacientes, se decidió estudiar el valor pronóstico de las cinco subunidades más recurrentemente alteradas en cáncer (*ARID1A*, *ARID1B*, *ARID2*, *SMARCA4* y *PBRM1*) (Centore et al., 2020) en LLA-B. Utilizando los datos de expresión y los datos clínicos de la misma cohorte de pacientes pediátricos con LLA-B y LLA-T del Capítulo III, se realizó un análisis de Cox univariable y multivariable de las covariables clínicas disponibles. Además, también se añadieron al estudio aquellas relacionadas con la expresión génica de las subunidades del complejo SWI/SNF mencionadas anteriormente dicotomizadas en dos grupos, “Bajo” y “Alto”, según el valor de la mediana de expresión.

De acuerdo con el análisis Cox univariable, además de las covariables clínicas ya analizadas en el Capítulo III, únicamente la expresión génica de las subunidades *ARID1A*, *ARID2* y *SMARCA4* influyeron de manera estadísticamente significativa en la supervivencia de los pacientes. Los pacientes con mayor expresión de *ARID1A* (p-valor = 0,027; CR: 0,1,870; IC del 95%: 1,073-3,261) y *SMARCA4* (p-valor = 0,012; CR: 2,039; IC del 95%: 1,168-3,560) estaban significativamente asociados con una supervivencia menor, mientras que una mayor expresión génica de *ARID2* se encontraba asociada con tiempos de supervivencia globales más largos (p-valor = 0,006; CR: 0,457; IC del 95%: 0,262-0,799) [**Figura 16B; Tabla S3**]. Tras realizar el análisis Cox multivariable con las covariables clínicas que presentaban un p-valor < 0,2 en el análisis univariable, únicamente la presencia de la fusión génica *TCF3-PBX1* resultó estar asociada significativamente con un mayor riesgo de muerte (p-valor =  $6,85 \times 10^{-4}$ ; CR: 9,401; IC del 95%: 2,579-34,274). Aun cuando ninguna de las covariables relacionadas con la expresión de las subunidades seleccionadas mostró resultados significativos mediante el análisis Cox multivariable, la subunidad *ARID2* fue la que mantuvo una significancia superior entre dichas subunidades (p-valor = 0,180; CR: 0,600; IC del 95%: 0,284-1,267) [**Figura 16A; Tabla S3**].

A diferencia de lo observado para la LLA-B, en el caso de la LLA-T no se encontró ninguna asociación entre la expresión de *ARID2* y la supervivencia del paciente (p-valor = 0,85) [**Figura 16B**], razón por la cual se decidió excluir este subtipo tumoral en los consiguientes estudios que se realizaron en modelos celulares *in vitro*. Dada la importancia funcional de *ARID2* en la formación de

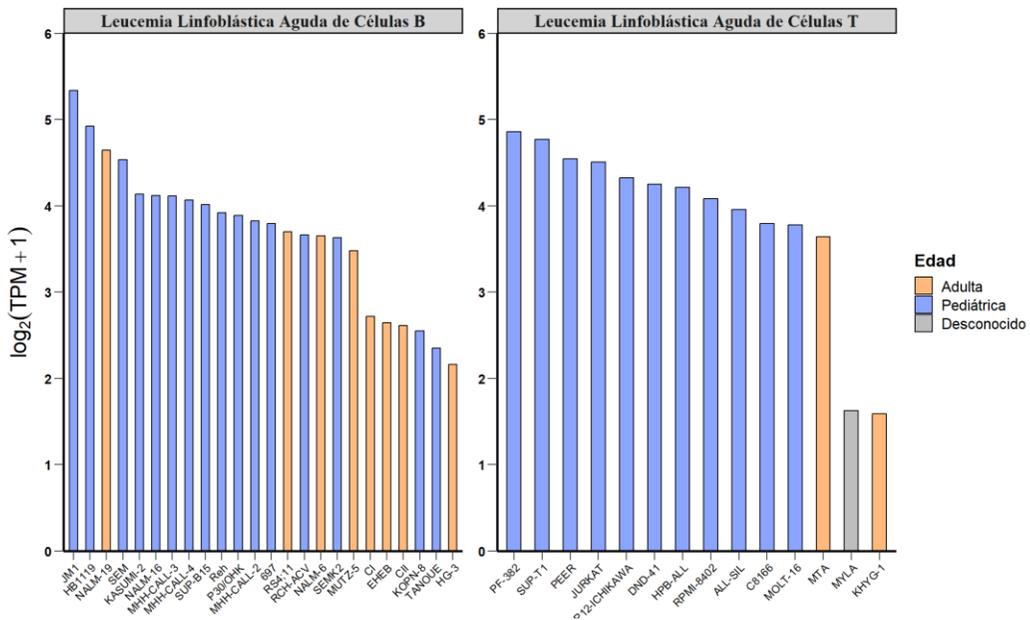
los complejos remodeladores de la cromatina PBAF, el bajo nivel de conocimiento de su implicación en LLA-B, y su mayor significancia en el análisis Cox multivariable, se decidió analizar el papel de esta subunidad en el contexto tumoral de la LLA-B.

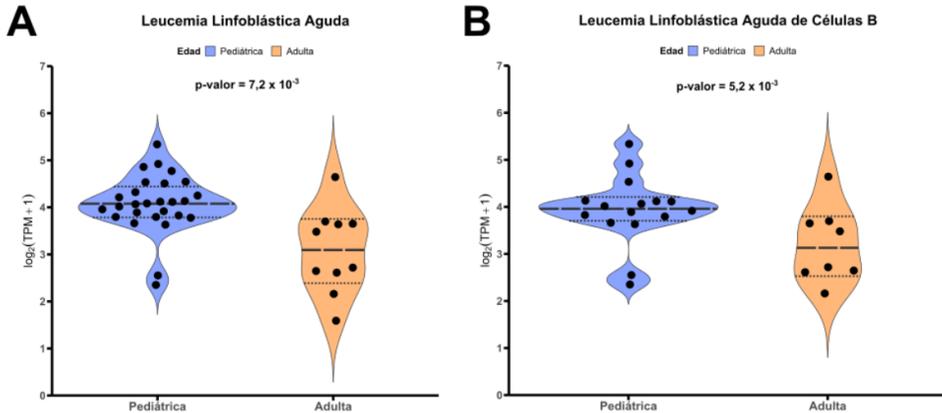


**Figura 16.** (A) Diagrama de Forest de las covariables clínicas seleccionadas en el análisis de Cox univariable ( $p$ -valor  $< 0,2$ ). Se representa el cociente de riesgo (CR) para cada covariable con su intervalo de confianza (IC) al 95%, junto al  $p$ -valor del análisis de Cox multivariable. La línea vertical punteada negra representa el valor de referencia 1 (sin riesgo de evento). (B) Curvas de supervivencia global Kaplan-Meier de los pacientes pediátricos con LLA-B ( $N = 120$ ) (izquierda) y LLA-T ( $N = 200$ ) (derecha) de acuerdo con la división de los grupos de expresión de los genes *ARID2* en “Bajo” y “Alto”. Se muestran los  $p$ -valores asociados a los análisis de supervivencia.

## 4.2.2 Niveles de expresión génica de *ARID2* en líneas celulares de LLA pediátricas

Tras analizar los niveles de expresión de *ARID2* en pacientes con LLA pediátricos, se descargaron y analizaron los datos de expresión génica de todas las líneas celulares de LLA disponibles procedentes de la base de datos Depmap en formato TPM [Figura 17]. No se observaron diferencias estadísticamente significativas en la expresión de *ARID2* entre las líneas celulares de LLA de células B y las de células T (p-valor = 0,582) [Figura 18A]. Alternativamente, tras analizar los datos según la edad del paciente del cual se obtuvo la línea celular de LLA, se determinó que la expresión de *ARID2* fue significativamente superior en las líneas de LLA pediátricas en comparación con las adultas (p-valor =  $7,2 \times 10^{-3}$ ) [Figura 18B]. De la misma manera, cuando este análisis se realizó independientemente según el subtipo celular, la diferencia significativa se mantuvo para las líneas celulares pediátricas de células B (p-valor =  $5,2 \times 10^{-3}$ ). No se pudo realizar una comparación robusta según la edad con las líneas celulares de LLA-T dado que solo existían dos catalogadas como adultas.





**Figura 18.** Comparación de la expresión génica de *ARID2* ( $\log_2(\text{TPM} + 1)$ ) de las líneas celulares LLA disponibles en la base de datos Depmap según la edad, pediátrica y adulta. **(A)** Valores correspondientes a las líneas celulares de LLA de subtipo B y T. **(B)** Valores correspondientes a las líneas celulares únicamente de subtipo B.

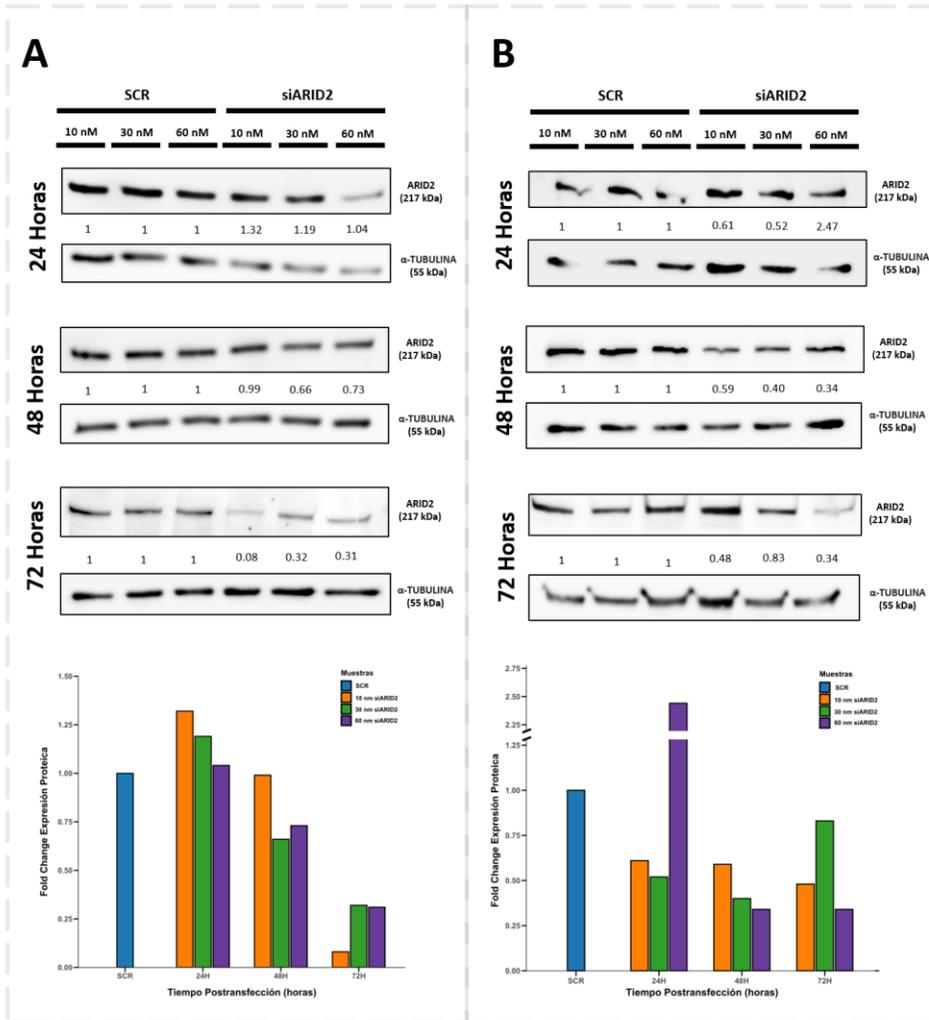
Existiendo diferencias significativas en la expresión de *ARID2* según la edad de la línea celular, se decidió utilizar las líneas celulares de LLA-B adulta RS4;11 y pediátrica SUP-B15 disponibles en el laboratorio para explorar los efectos fenotípicos de la modificación de la expresión de esta subunidad en el contexto tumoral de la leucemia.

### 4.2.3 Optimización del silenciamiento génico de *ARID2* en líneas celulares de LLA-B

Para analizar si el cambio de expresión de *ARID2* posee un efecto fenotípico celular, se procedió a realizar su silenciamiento con el uso de tecnología de ARNsi en las líneas celulares de LLA-B RS4;11 y SUP-B15. En primer lugar, se llevó a cabo una optimización de la concentración de ARNsi de *ARID2* (10 nM, 30 nM, 60 nM) y el tiempo de postransfección (24 horas, 48 horas, 72 horas) necesario para conseguir la mayor inhibición posible de su expresión proteica monitorizada mediante Western blot.

En el caso de la línea celular LLA-B adulta RS4;11, se obtuvieron niveles altos de silenciamiento de *ARID2* transcurridas 72 horas tras la transfección, alcanzándose un 92% de reducción de la expresión con una concentración de 10 nM del ARNsi [Figura 19A]. Para la línea celular pediátrica SUP-B15, aunque el silenciamiento no fue tan efectivo como en RS4;11, se llegó a reducir un 66% la expresión proteica de *ARID2* con 60 nM del ARNsi 72 horas tras la transfección. Sin embargo, dado que a las 48 horas se conseguía un porcentaje de inhibición

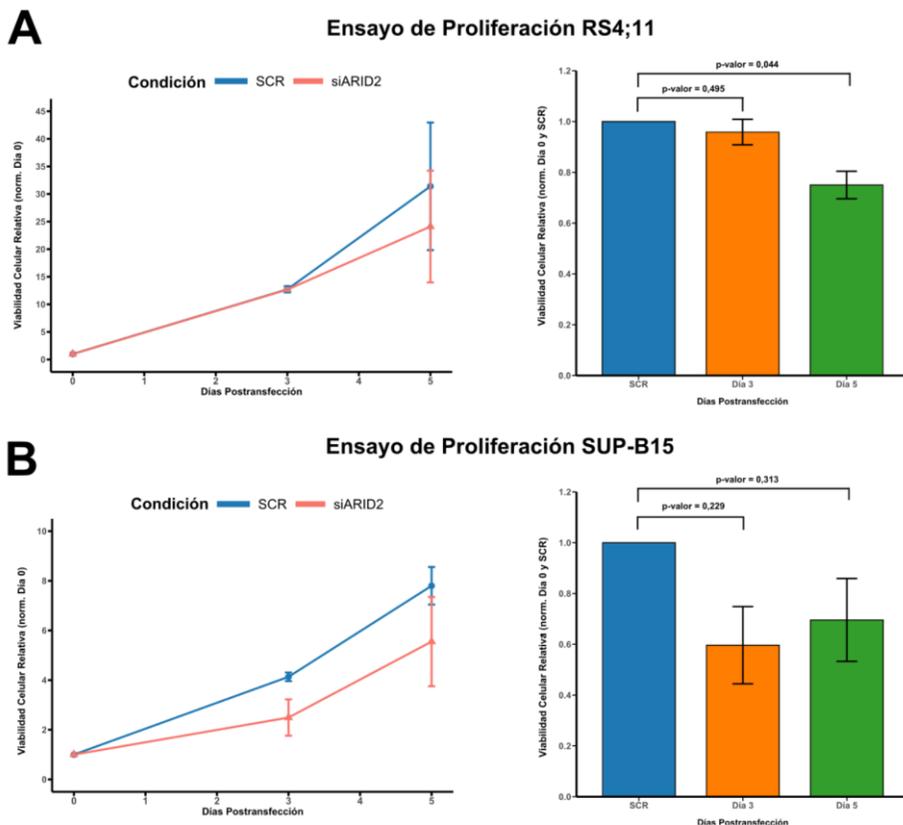
ligeramente inferior del 60% usando la mitad de concentración del ARNsi, se decidió utilizar estas condiciones para los análisis posteriores [Figura 19B].



#### 4.2.4 El silenciamiento de *ARID2* disminuye la viabilidad y proliferación celular

Para cuantificar la viabilidad celular de las células transfectadas RS4;11 (10 nM, 72 horas) y SUP-B15 (30 nM, 48 horas) con SCR y siARID2, se utilizó el ensayo de resazurina.

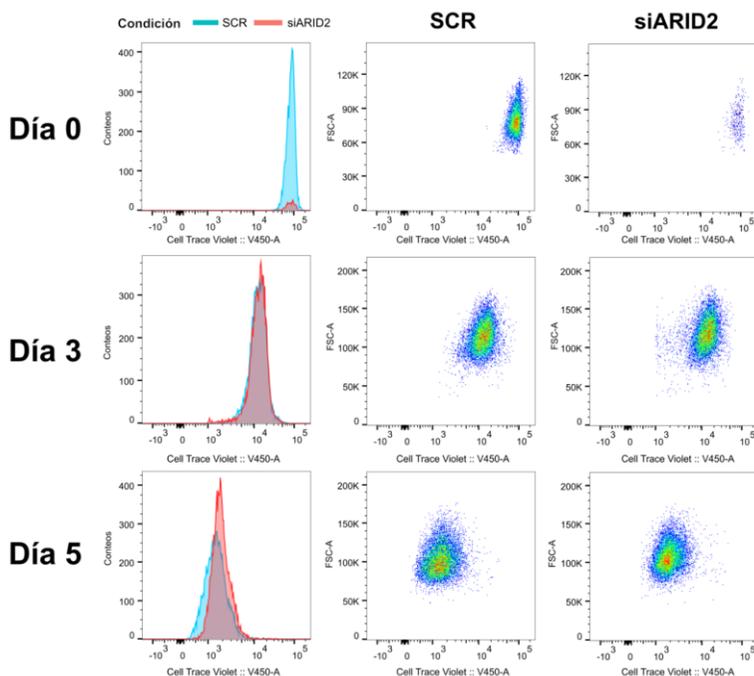
En el caso de la línea celular de LLA-B adulta RS4;11 tratada con siARID2, no se evidenció ningún cambio en la viabilidad celular durante los tres primeros días tras la transfección en comparación con la condición control SCR (p-valor = 0,495). En cambio, se obtuvo una media de disminución de viabilidad significativa del 25% en el día 5 (p-valor = 0,044) [Figura 20A].



**Figura 20.** Ensayos de proliferación realizados con resazurina para las líneas celulares RS4;11 (N = 3) (A) y SUP-B15 (N = 2) (B) con las condiciones optimizadas de transfección: [10 nM, 72 horas] y [30 nM, 48 horas], respectivamente. A la izquierda, se muestra la evolución de la proliferación a días 3 y 5, normalizando con día 0. A la derecha, el diagrama de barras muestra la viabilidad celular de la condición siARID2 respecto al día 0 y a la condición negativa “scramble” (SCR).

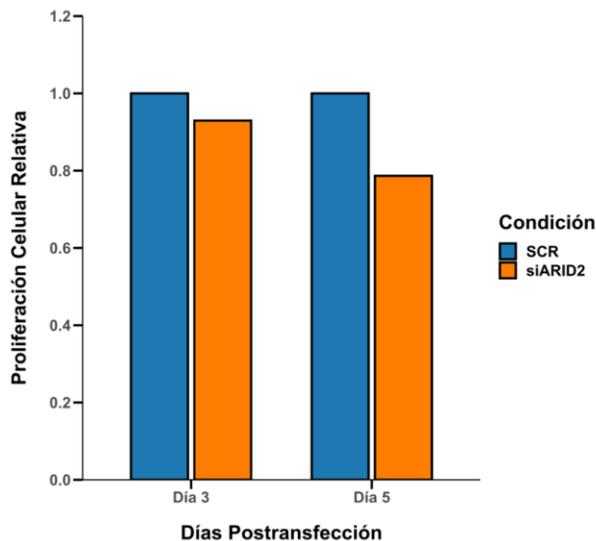
Aunque la línea celular de LLA-B pediátrica SUP-B15 tratada mostró una misma tendencia en la disminución de su viabilidad celular, no se observaron diferencias estadísticamente significativas para los días 3 (p-valor = 0,229) y 5 (p-valor = 0,313). A pesar de esto, la media de la reducción de la viabilidad observada en ambos días fue del 40% y 30%, respectivamente. [Figura 20B].

En base a los resultados en el ensayo de viabilidad, para saber si el silenciamiento de *ARID2* estaba afectando a la tasa de proliferación celular o a la muerte celular, se realizó un ensayo de proliferación con el kit CellTrace™ Violet mediante citometría de flujo con la línea celular de LLA-B adulta con los mismos puntos de medida a día 3 y día 5. En este ensayo, el desplazamiento de la curva hacia la izquierda, representando una menor intensidad de fluorescencia detectada, indica una mayor proliferación celular. Así, se pudo visualizar la disminución de la intensidad de fluorescencia del colorante con el paso del tiempo, obteniéndose una mayor diferencia de bajada de la señal a día 5 con el SCR. En este punto temporal, la menor tasa de proliferación de la línea celular tratada con siARID2 concuerda con el resultado obtenido en el ensayo de viabilidad, donde tampoco existen diferencias a día 3 [Figura 21].



**Figura 21.** Citometría de flujo de la línea celular de LLA-B adulta RS4;11 marcada con el tinte CellTrace™ Violet. Se muestra la evolución de la intensidad de fluorescencia de la condición negativa “scramble” (SCR) y la tratada con siARID2.

Realizando una normalización de cada condición siARID2 con el SCR en cada uno de los puntos temporales, a día 3 la disminución de la proliferación inapreciable visualmente fue del 7,05%, mientras que a día 5 llegó hasta un 21,33%. Sin embargo, se debe tener en cuenta que los resultados hacen referencia a un único experimento, por lo que en el futuro debe confirmarse con un mayor número de réplicas biológicas para confirmar el resultado [Figura 22].



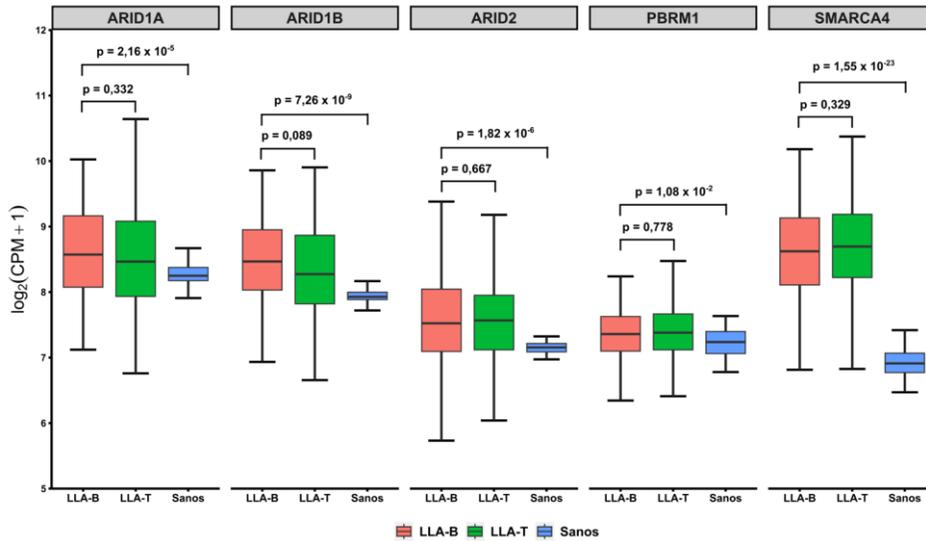
**Figura 22.** Proliferación celular relativa de RS4;11 tratada con siARID2 en comparación con la condición negativa “scramble” (SCR).

#### 4.2.5 Menor expresión génica de ARID2 en pacientes pediátricos sanos.

Durante los experimentos *in vitro* explicados en las anteriores secciones, nuevos datos transcriptómicos de ARN-Seq de pacientes sanos del proyecto TARGET-AML fueron incorporados al portal GDC. Sabiendo que las muestras relativas a estos datos no eran pareadas de pacientes con LMA, se consideraron controles sanos para comparar con los datos transcriptómicos de pacientes con LLA-B y con LLA-T pediátricos de nuestro estudio. Así, se compararon los datos de expresión génica de las subunidades ARID1A, ARID1B, ARID2, SMARCA4 y PBRM1 en formato  $\log_2(\text{CPM} + 1)$  de todos estos pacientes con la cohorte de TARGET-LLA estudiada para conocer cómo varían los niveles en el contexto tumoral.

De acuerdo con la comparativa realizada, se obtuvo una mayor expresión estadísticamente significativa de los pacientes con LLA-B pediátricos en comparación con las muestras sanas en todas las subunidades. En este caso, la mayor

sobre-expresión se obtuvo para la subunidad SMARCA4 ( $p$ -valor =  $1,55 \times 10^{-23}$ ). De esta manera, se pudo confirmar que existía una expresión génica aberrante de *ARID2* en pacientes LLA-B pediátricos en comparación con pacientes sanos ( $p$ -valor =  $1,82 \times 10^{-4}$ ). Sin embargo, realizando la comparación con los pacientes con LLA-T pediátricos, ninguna de las subunidades estudiada se encontraba diferencialmente expresada de manera significativa [Figura 23].



**Figura 23.** Comparación de la expresión génica de *ARID2* ( $\log_2(\text{CPM} + 1)$ ) entre las muestras de LLA-B (N = 120), LLA-T (N = 200) y Sanos (N = 50) derivadas de los pacientes pediátricos del proyecto TARGET(A). Se muestran los p-valores correspondientes a LLA-B vs LLA-T y LLA-B vs Sanos.

### 4.3 Discusión

A pesar de categorizar los genes en dos grupos opuestos según su papel en el desarrollo tumoral, oncogenes y genes supresores tumorales, actualmente se ha demostrado que algunos de ellos pueden exhibir ambas funciones en distintos contextos celulares (L.-H. Wang et al., 2018). El papel dual que pueden ejercer estos genes se encuentra justificado tanto por la compleja red de señalización y los diferentes niveles de regulación que coexisten a nivel intracelular, como por el perfil genético específico de cada tipo celular (Lou et al., 2014). Se ha observado que la ruta de señalización de Notch, que posee un papel crucial en la regulación de la diferenciación y determinación celular del sistema hematopoyético, puede actuar como supresora tumoral en algunas neoplasias hematológicas de la línea mieloide, mientras que se ha caracterizado en varios estudios su implicación oncogénica en enfermedades relacionadas con la línea linfóide (Grabher et al., 2006; Lobry et al., 2013). La fosfatasa PTEN se encuentra deletada frecuentemente en la LLA-T asociándose con la renovación de las células madre leucémicas, mientras que su pérdida de función en la LLA-B inhibe la proliferación celular (Martelli et al., 2019; Shojaee et al., 2010). De esta manera, al contrario que en la LLA-B, la alta expresión de PTEN es un factor pronóstico favorable en los pacientes con LLA-T (Datta et al., 2020). Así pues, se debe tener en cuenta que una terapia génica dirigida a un gen específico con esta característica dual puede ser efectiva para determinadas líneas celulares tumorales, pero podría promover la supervivencia en otros tipos celulares.

Algunas subunidades del complejo SWI/SNF han sido descritas en la bibliografía por su papel dual como supresor tumoral y como factores implicados en el mantenimiento de tumores. En el caso de *ARID1A*, conocido por ser un gen supresor de tumores, se ha observado que su papel en la oncogénesis depende del contexto tisular (Otto & Kadoch, 2017). Por ejemplo, la expresión de *ARID1A* es crucial para el inicio del hepatocarcinoma celular, pero su pérdida en homocigosis o heterocigosis en tumores avanzados acelera su progresión (Sun et al., 2017). De igual manera, se ha descrito que *SMARCA4* actúa como supresor tumoral en diversos tipos de cáncer, tales como en cáncer de pulmón (Fillmore et al., 2015) o en cáncer de ovario (Witkowski et al., 2014). En cambio, se ha observado que está involucrado en la regulación de rutas que promueven la oncogénesis en otros tipos de cáncer, por ejemplo, en cáncer de próstata (Giles et al., 2021). Específicamente en las enfermedades hematológicas, por ejemplo, se ha observado que *BCL11B* puede actuar como supresor tumoral en LMC, dado que su haploinsuficiencia junto a la presencia del gen de fusión en *BCR-ABL* contribuyen al desarrollo de la enfermedad (Nagamachi et al., 2009). Sin embargo, en las leucemias agudas de linaje ambiguo, *BCL11B* se encuentra activado de manera aberrante en

las HSC, produciendo un bloqueo de la diferenciación mieloide y activando inadecuadamente los genes implicados en la diferenciación del linaje celular T (Montefiori et al., 2021). Por otra parte, en otras neoplasias, es necesaria la función fisiológica de la subunidad para mantener el desarrollo tumoral. En el caso de *BCL11A*, se ha observado que su pérdida de función en el linfoma de células T/*Natural Killer* reduce la proliferación de las células tumorales (Shi et al., 2020). Por tanto, la implicación del complejo SWI/SNF en el desarrollo tumoral es heterogénea dependiendo del contexto tisular en el que se encuentre, así como de la accesibilidad característica de la cromatina en cada tipo celular (Corces et al., 2018).

De acuerdo con el último estudio ómico realizado sobre una de las mayores cohortes de pacientes pediátricos con LLA ( $N = 2.754$ ), se propuso a *ARID2* como posible candidato gen conductor en LLA-B. *ARID2* se encontraba alterado genéticamente en al menos un 2,5% de los pacientes pediátricos con LLA-B ( $N_T = 1.428$ ) (Brady et al., 2022), mientras que en estudios pasados se detectó mutado en aproximadamente el 5% de los pacientes de una cohorte más pequeña ( $N_T = 218$ ) (X. Ma et al., 2018). En los resultados presentados en este Capítulo IV relacionados con los pacientes con LLA-B pediátricos procedentes del proyecto TARGET, los niveles de ARNm de *ARID2* se encontraban sobre-expresados de manera significativa en comparación con las muestras de médula ósea sanas disponibles. No obstante, según los datos disponibles sobre mutaciones en esta cohorte en el portal GDC, no existía ninguna mutación en *ARID2* que pudiera correlacionarse con la expresión observada. Por ello, en este Capítulo IV, no se ha podido demostrar la causa genética que incrementa de manera aberrante la expresión de *ARID2* en muestras de LLA-B pediátricas. Aunque *ARID2* puede ser considerado supresor tumoral en otras enfermedades, los resultados señalan que en la LLA-B pediátrica podría tener un rol diferente como ocurre de manera similar con otras subunidades del complejo SWI/SNF.

Después del comienzo de esta investigación, Bluem et al. propusieron que la pérdida de *ARID2* potenciaba el desarrollo tumoral en los estadios iniciales en modelos de LMA *in vivo*, aunque su funcionalidad es necesaria para el mantenimiento de esta neoplasia. De esta manera, en los primeros estadios *ARID2* es considerado un posible supresor tumoral, mientras que, una vez desarrollada la leucemia, es considerado oncogen para el progreso tumorigénico (Bluemn et al., 2022). Los resultados obtenidos en este Capítulo IV para la subunidad *ARID2* en LLA-B parecen apoyar la hipótesis de Bluem et al. en LMA con respecto al posible rol dual de *ARID2* en tumorigénesis. Si bien la sobreexpresión de *ARID2* podría estar correlacionada con una mayor supervivencia global de los pacientes, lo que sugeriría un papel supresor tumoral, al mismo tiempo se encuentra sobre-

expresado en LLA-B en comparación con muestras sanas. Además, los ensayos *in vitro* parecen sugerir un rol oncogénico ya que su silenciamiento disminuye la viabilidad y proliferación celulares. Sin embargo, la expresión de ninguna de las subunidades del complejo SWI/SNF propuestas influyó de manera significativa en el riesgo de muerte según el análisis Cox multivariable para considerarse factor pronóstico independiente en la LLA-B pediátrica. Una limitación de este estudio es la falta de la correlación de la expresión de estas subunidades con la supervivencia libre de enfermedad, ya que esta podría estar asociada con una mejor o peor respuesta al tratamiento. Por ejemplo, se ha demostrado que una baja expresión de MN1 se encuentra asociada a una mejor respuesta al tratamiento en pacientes mayores con LMA con características citogenéticas normales (Schwind et al., 2011).

Por otra parte, no existen modelos *in vitro* que caractericen la importancia de *ARID2* en el contexto de la LLA-B. En los resultados presentados, relacionados con la inhibición transitoria de *ARID2* en las líneas celulares de LLA-B de RS4;11 y SUP-B15, se ha demostrado por primera vez la posible implicación de esta subunidad en la regulación de la proliferación celular de la línea linfocítica de células B. Aunque se requiere un mayor número de réplicas biológicas para corroborar la disminución de viabilidad celular observada, los resultados apuntan a que una bajada de la expresión de *ARID2* en LLA-B podría afectar a la proliferación de las células leucémicas. Según los resultados obtenidos en los ensayos de viabilidad, donde la línea celular de LLA-B pediátrica SUP-B15 presentó la mayor reducción, posiblemente el efecto en la proliferación en ensayos futuros sea más drástico en comparación con la línea celular adulta RS4;11. Según los resultados obtenidos en la comparación realizada con datos de expresión procedentes de DepMap, el mayor efecto en la viabilidad observado en SUP-B15 podría estar justificado por la mayor sobre-expresión de *ARID2* significativa que poseen los modelos de LLA-B pediátricos en comparación con los modelos adultos. En conjunto con los resultados anteriores relacionados con los pacientes pediátricos, se refuerza la posibilidad de que *ARID2* puede estar implicado en el mantenimiento tumoral de la LLA-B.

La eliminación de *ARID2*, tal como se mencionó en la Introducción en relación con su relevancia estructural en los complejos remodeladores de la cromatina, supone la pérdida de conformación del complejo PBAF y la desregulación de programas transcripcionales específicos tanto en los estadios iniciales de la enfermedad como en los modelos de mantenimiento leucémicos mieloides estudiados (Bluemn et al., 2022). A nivel genético, se ha propuesto que el efecto obtenido en las células madre debido a la eliminación de *ARID2* se encuentra relacionado con una regulación al alza de las vías inflamatorias (Bluemn et al.,

2021). Más recientemente, el mismo autor demostró que, para los estadios iniciales, la eliminación de ARID2 supuso una desregulación de la expresión génica de los factores de transcripción GATA2, E2-2 y SIX5 y, por tanto, de los programas transcripcionales que controlan aguas abajo. Sin embargo, ninguno de estos genes estuvo presente en los expresados diferencialmente de manera significativa en los modelos relacionados con el papel oncogénico de *ARID2*, en el cual destacaba la presencia de un factor de transcripción de la familia Ets denominado Etv5. Dado que Etv5 se trata de un oncogen caracterizado en la LMA interactor directo con Arid2, los autores propusieron que posiblemente medie el fenotipo tumoral en el modelo de mantenimiento de LMA estudiado (Bluemn et al., 2022; Takeda et al., 2020). En la línea de investigación de este Capítulo IV, futuros experimentos de secuenciación ARN-Seq entre de las líneas celulares de LLA-B controles y tratadas por el siARID2 permitirán dilucidar las vías biológicas que regula ARID2 de manera independiente o conjunta con el complejo PBAF en la LLA-B.

# Capítulo V

## Implicación de ARID1A y ARID1B en cáncer

# CAPÍTULO V: Implicación de ARID1A y ARID1B en cáncer

El Capítulo V muestra los resultados obtenidos relacionados con el Objetivo 2.2, enfocado en el análisis del fenotipo que produce la delección de *Arid1a* y *Arid1b* en modelos *in vitro* e *in vivo* de ratón. Mediante el uso de la transcriptómica y la medición de distintos parámetros biológicos, se pretende conocer en mayor profundidad el impacto biológico de las alteraciones de estas subunidades del complejo remodelador de la cromatina SWI/SNF en la carcinogénesis.

## 5.1 Antecedentes

### 5.1.1 Complejos SWI/SNF residuales en el desarrollo tumoral

Como se ha mencionado anteriormente, la alta frecuencia mutacional del complejo SWI/SNF en diversos tipos de cáncer indica que el control que ejerce en la regulación epigenética juega un papel fundamental en el desarrollo tumoral (Tolstorukov et al., 2013). Las mutaciones presentes en los genes de las subunidades del complejo SWI/SNF a menudo generan dependencias específicas por otras subunidades. En consecuencia, estas mutaciones no desactivan completamente al complejo SWI/SNF, sino que promueven el desarrollo tumoral a través de la dependencia de la célula por otros complejos SWI/SNF residuales alternativos (Hohmann & Vakoc, 2014). Por ejemplo, aunque las alteraciones genéticas de *SMARCB1* en tumores rabdoideos generen una pérdida de función, existen complejos SWI/SNF residuales con BRD9 necesarios para la supervivencia de las células tumorales (X. Wang et al., 2019). Este hallazgo plantea la posibilidad de que los cánceres con mutaciones en *SMARCB1* no son impulsados únicamente por la ausencia de esta subunidad, sino más bien por un funcionamiento anormal de los complejos remodeladores de la cromatina disponibles en la célula (Mittal & Roberts, 2020).

Actualmente, existe un conocimiento limitado sobre cómo los complejos SWI/SNF defectuosos contribuyen al desarrollo del cáncer debido, principalmente, a la similitud en la composición estructural que presentan las distintas familias conocidas. Por ejemplo, las subunidades ATPasa SMARCA2 y SMARCA4 se pueden encontrar en los tres complejos caracterizados de la familia SWI/SNF:

cBAF, PBAF y ncBAF (Centore et al., 2020). De este modo, la eliminación de una de estas subunidades dificulta el desensamblaje dirigido de manera específica a uno de los tres complejos sin poder evitar la posibilidad de provocar efectos secundarios en el resto. Adicionalmente, las funciones que pueden ejercer una misma subunidad y su paróloga pueden variar considerablemente entre distintos tipos de tejidos, lo cual aumenta la heterogeneidad del posible impacto en el desarrollo tumoral (Reisman et al., 2005).

### **5.1.2 Letalidad sintética de subunidades del complejo SWI/SNF**

En cáncer, la letalidad sintética se define como un evento donde la alteración genética simultánea de múltiples genes provocan la muerte celular, mientras que si ocurre de manera individualizada resulta compatible con la supervivencia de la célula tumoral. Existen diversos factores que pueden afectar al mecanismo de letalidad sintética, tales como la regulación epigenética, las rutas metabólicas o el microambiente tumoral específico (Topatana et al., 2020).

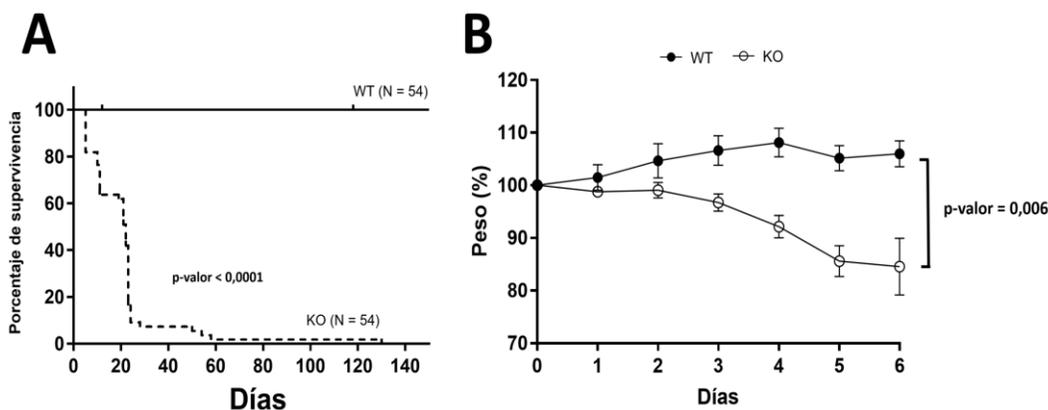
En este contexto, el papel de supresor tumoral de *ARID1A* y su alta frecuencia mutacional demostrado en diversos estudios abre una vía de investigación sobre sus posibles asociados letales sintéticos para la búsqueda de dianas terapéuticas. Las líneas celulares mutantes para la subunidad ARID1A muestran una dependencia específica de la subunidad paróloga ARID1B (Helming et al., 2014). Sin embargo, aún no se ha estudiado en profundidad los efectos fenotípicos que resultan de la alteración de ambas subunidades parálogas en modelo *in vivo* ni la dependencia que se puede generar por otros complejos remodeladores de la cromatina residuales.

## 5.2 Resultados

### 5.2.1 Parámetros biológicos afectados en los ratones DKO *Arid1a/Arid1b*

Inicialmente, se realizó una comparación de la supervivencia entre los ratones DKO (*Arid1a*<sup>-/-</sup>*Arid1b*<sup>-/-</sup>) y los modelos WT (*Arid1a*<sup>+/+</sup>*Arid1b*<sup>+/+</sup>). El resultado de la delección de los genes *Arid1a* y *Arid1b* tras la inyección de tamoxifeno redujo drásticamente la supervivencia del modelo DKO (p-valor < 0,0001) [Figura 24A]. La menor supervivencia de estos ratones con los genes *Arid1* delecionados condujo a la exploración de los parámetros biológicos alterados en esos ratones.

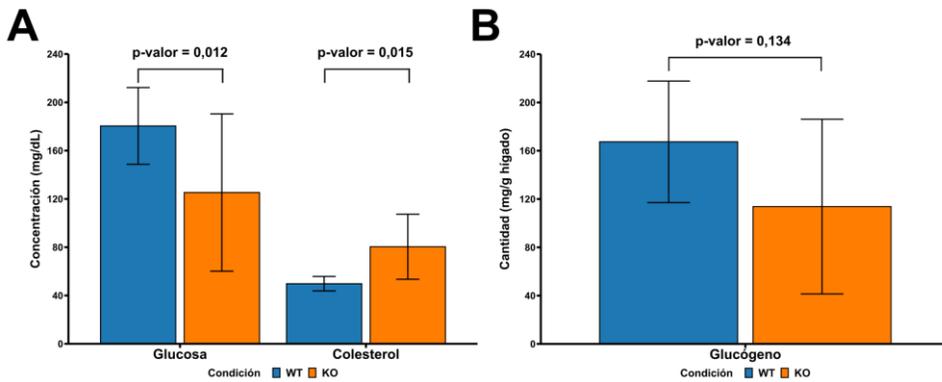
En primer lugar, se pudo observar que los ratones DKO mostraban una disminución del peso a lo largo de los días tras el tratamiento con tamoxifeno. La mayor reducción de peso se obtuvo en el día 6 siendo hasta un 21% menor que los ratones WT (p-valor = 0,006) [Figura 24B].



**Figura 24.** (A) Curvas de supervivencia global Kaplan-Meier de los ratones DKO *Arid1a/Arid1b* (N = 54) y de los WT (N = 54). (B) Porcentaje de cambio de peso de los ratones DKO *Arid1a/Arid1b* medidos 6 días después tras de la inyección de tamoxifeno.

Adicionalmente, se recogieron datos sobre distintos parámetros hematológicos y serológicos de los ratones con el fin de analizar las consecuencias de la delección de los genes *Arid1*. Entre los hallazgos más relevantes, se encuentra una reducción significativa de los niveles de glucosa (p-valor = 0,012) y un aumento significativo de la cantidad de colesterol (p-valor = 0,015) en sangre en el modelo DKO [Figura 25A]. Por otro lado, aunque las diferencias no fueron estadísticamente significativas, se observó una reducción de los niveles de glucógeno hepático en el modelo DKO (p-valor = 0,134) [Figura 25B]. Los parámetros biológicos restantes relacionados con la hematología así como las diversas enzimas séri-

cas asociadas a los tejidos analizados, no presentaron diferencias significativas (información no mostrada).

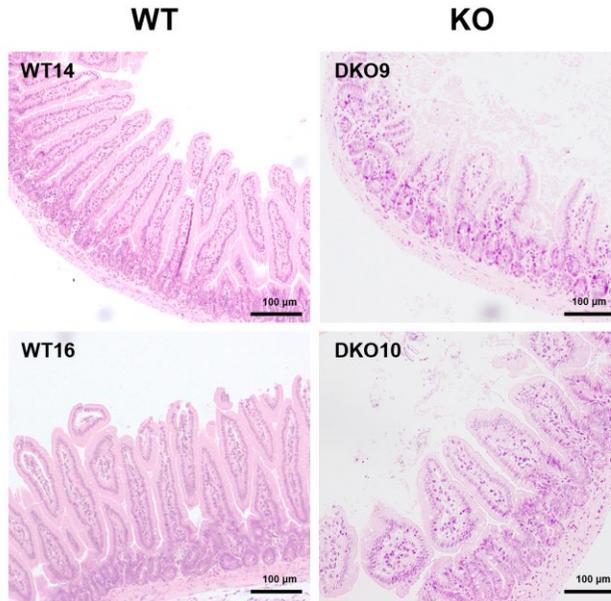


**Figura 25.** Comparación de los niveles de glucosa y colesterol en sangre **(A)**, y la cantidad de glucógeno en el hígado **(B)** entre ratones DKO *Arid1a/Arid1b* y los WT. El número de réplicas (DKO vs WT) comparadas fueron: Glucosa (11 vs 14), colesterol (6 vs 10) y glucógeno (6 vs 10).

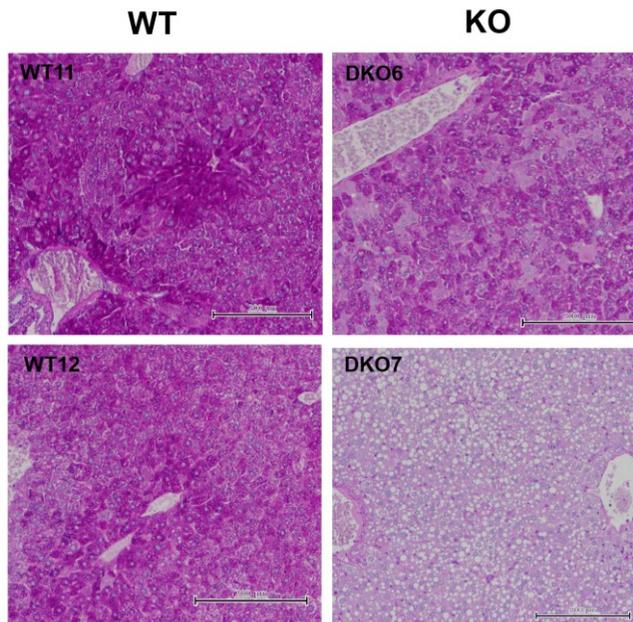
### 5.2.2 Análisis histológico de los ratones DKO *Arid1a/Arid1b*

Para explorar el grado de afectación de los órganos del modelo DKO de los genes *Arid1*, se decidió analizar histológicamente diferentes tejidos mediante tinciones de hematoxilina-eosina con el objetivo de encontrar una explicación fisiopatológica al fenotipo drástico observado. Debido a la acusada pérdida de peso y alteraciones metabólicas de los ratones DKO, la primera hipótesis planteada fue que la doble delección de *Arid1a/Arid1b* podía estar afectando a la estructura del aparato digestivo y la absorción de nutrientes. Tras analizar los cortes del intestino delgado, se observó efectivamente la desorganización y el acortamiento de la estructura de las vellosidades del intestino delgado en los modelos DKO en comparación con los WT **[Figura 26]**.

Dado el nivel de reducción de glucógeno mencionado en la anterior sección, se decidió realizar la tinción de PAS para analizar la presencia de los niveles de estos polisacáridos en el hígado. El colorante de la tinción PAS torna a un color rojizo en presencia de grupos aldehído formados por oxidación previa de carbohidratos. Como resultado, se pudo observar que el modelo DKO presentaba una menor intensidad de coloración rojiza en las secciones de hígado en comparación con los ratones WT. Adicionalmente, la muestra DKO7 presentaba visualmente acumulaciones de grasas características de la esteatosis hepática **[Figura 27]**.

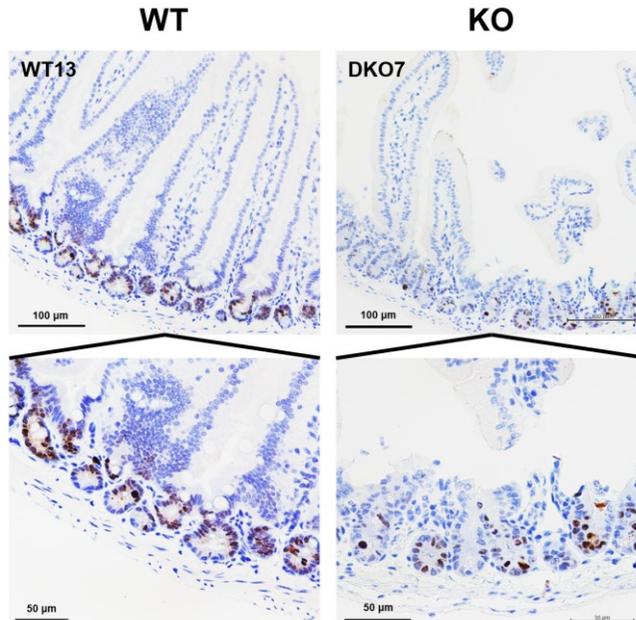


**Figura 26.** Tinción hematoxilina-eosina de la arquitectura de las vellosidades intestinales en muestras de yeyuno de ratones WT (izquierda) y DKO *Arid1a/Arid1b* (derecha). Se muestran los ratones WT14, WT16, DKO9 y DKO10.



**Figura 27.** Tinción PAS para la visualización de glucógeno en muestras de hígado de ratones WT (izquierda) y DKO *Arid1a/Arid1b* (derecha). Se muestran los ratones WT11, WT12, DKO6 y DKO7.

La delección de *Arid1a* y *Arid1b* ha sido recientemente vinculada también con una alta tasa de proliferación celular y desdiferenciación en algunos tejidos, por ejemplo, en hígado (Z. Wang et al., 2020). En relación con el intestino delgado, Hiramatsu et al. no obtuvieron diferencias significativas en la proliferación celular de las criptas intestinales entre ratones KO *Arid1a* específicos de este tejido y el modelo WT (Hiramatsu et al., 2019). Por ello, se decidió realizar una inmunohistoquímica de la proteína KI67 para estudiar si la capacidad proliferativa celular estaba afectada en alguno de los tejidos analizados previamente con mayores signos patológicos. En este caso, únicamente se obtuvieron resultados significativos al analizar las criptas de Lieberkühn presentes en las muestras de yeyuno. Se pudo apreciar visualmente que existía una disminución de proliferación en las muestras DKO en comparación con los ratones WT de manera consistente en todos los individuos analizados [Figura 28].

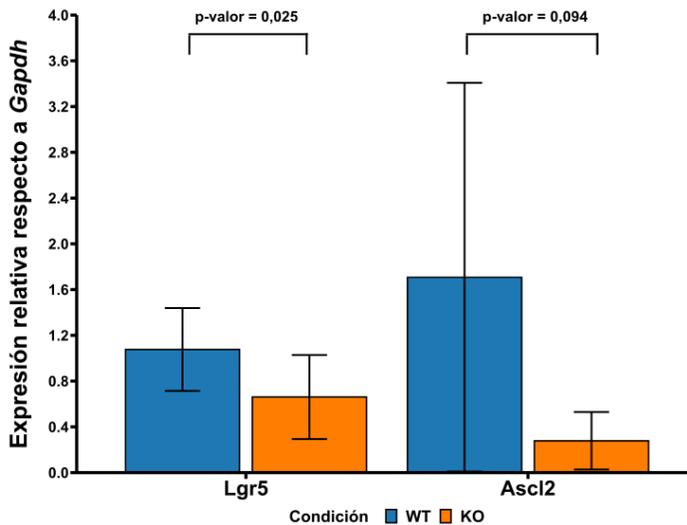


**Figura 28.** Inmunohistoquímica de *Ki67* para evaluar signos de proliferación en las criptas de Lieberkühn de las muestras de yeyuno de ratones WT (izquierda) y DKO *Arid1a/Arid1b* (derecha). Se muestran los ratones WT13 y DKO7. Las imágenes inferiores muestran un aumento de zoom de las imágenes superiores.

### 5.2.3 Menor expresión de marcadores de células madre intestinales en el DKO *Arid1a/Arid1b*

La estructura desorganizada de las criptas intestinales y la disminución de la proliferación observada en las inmunohistoquímicas realizadas con *Ki67* de

los modelos DKO *Arid1a/Arid1b* condujo al análisis de marcadores de células madre intestinales que pudiesen estar desregulados en estas estructuras de las vellosidades. Las proteínas LGR5 y ASCL2, entre otras, son marcadores característicos las células madre en cuestión involucradas en la homeostasis de la renovación celular en las criptas intestinales (Y. Liu et al., 2023). Así, para evaluar si existía una disminución de la capacidad proliferativa de estas células madre, se procedió a cuantificar la expresión génica de *Lgr5* y *Ascl2* en criptas intestinales purificadas de los ratones WT y DKO *Arid1a/Arid1b*. Como resultado, se obtuvo una disminución únicamente significativa de *Lgr5* (p-valor = 0,025), mientras que se apreció la misma tendencia en los modelos DKO para *Ascl2* (p-valor = 0,094) [Figura 29].



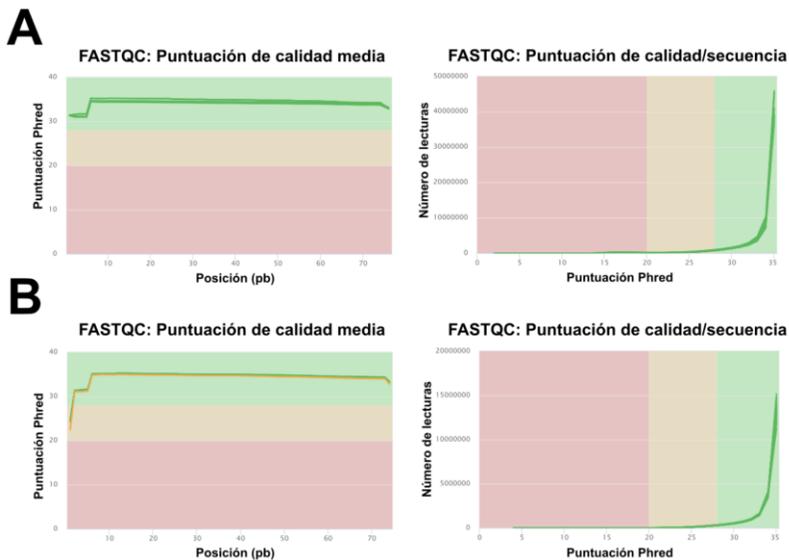
**Figura 29.** Análisis por qPCR de la expresión génica de los marcadores de célula madre intestinales *Lgr5* y *Ascl2* de ratones DKO *Arid1a/Arid1b* (N = 10) y los WT (N = 6). Se representa la media de expresión de la condición WT y KO normalizada con el gen constitutivo *Gapdh*. La desviación estándar representada deriva de muestras biológicas.

### 5.2.4 Control de calidad de los datos de secuenciación

Una vez evaluados los efectos fenotípicos de la delección de *Arid1a* y *Arid1b*, se decidió estudiar los cambios transcriptómicos que producían en los modelos DKO. Ambas subunidades forman parte del complejo cBAF, por lo que cabría esperar que su alteración pueda conducir a un cambio en la regulación de la expresión génica en la célula. Por ello, se realizó un ARN-Seq de 6 muestras

(3 WT vs. 3 DKO) de un modelo *in vitro* basado en cultivos primarios de MEF, y otro ARN-Seq de 10 muestras de tejidos de hígado y 10 muestras de tejidos de yeyuno más afectados anteriormente. Se seleccionaron los tejidos, según las características histológicas, de los DKO mayormente afectados a nivel microscópico: KO1, KO3, KO5, KO6, KO7. En el caso de los modelos WT, se seleccionaron las muestras WT11, WT12, WT13, WT14 y WT15.

Una vez obtenidos los ficheros de secuenciación crudos FASTQ procedentes de las dos secuenciaciones RNA-Seq realizadas, se ejecutó un análisis de calidad mediante la herramienta FASTQC. En general, la calidad de los archivos de secuenciación fue alta en cada una de las posiciones de las lecturas obtenidas (Escala Phred > 30) tanto en la secuenciación de las muestras de cultivos primarios de MEF [Figura 30A] como de los tejidos de ratón [Figura 30B].

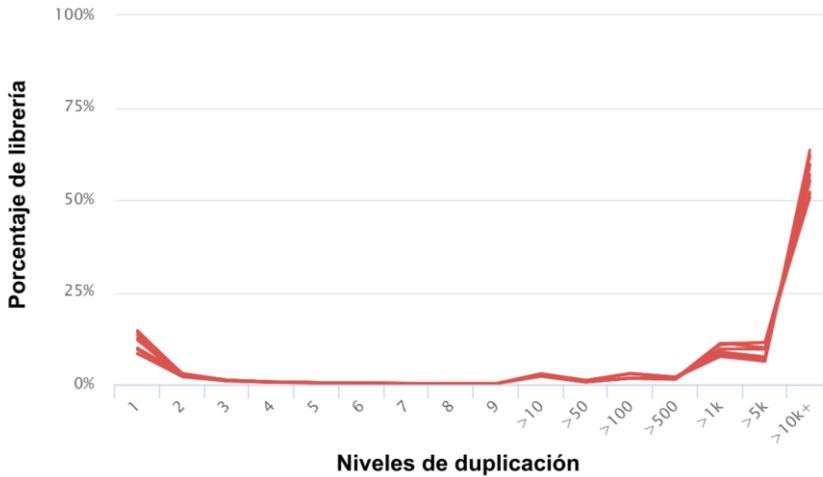


**Figura 30.** Análisis de calidad de los ficheros FASTQ derivados de secuenciación ARN-Seq realizado por la herramienta FASTQC en las muestras de MEF (A) y de tejido de ratón (B). A la izquierda, se muestra la puntuación de calidad media en cada una de las posiciones de las lecturas. A la derecha, se muestra la distribución de la calidad de todas las lecturas. La calidad se muestra en escala Phred.

Sin embargo, en el caso de las muestras derivadas de MEF, la herramienta FASTQC mostró un problema relacionado con la presencia de lecturas duplicadas en altos niveles [Figura 31]. En este punto, dado que no estaba relacionado con la presencia de adaptadores, se decidió explorar qué tipo de secuencias se encontraban recurrentemente duplicadas. Como resultado, se observó que hasta el 60% de las lecturas mapeadas lo hacían en genes relacionados con ARNr en todas las muestras de MEF. A pesar de ello, se decidió continuar estudiando los

genes diferencialmente expresados en los MEF DKO descartando las lecturas relacionadas con ARNr.

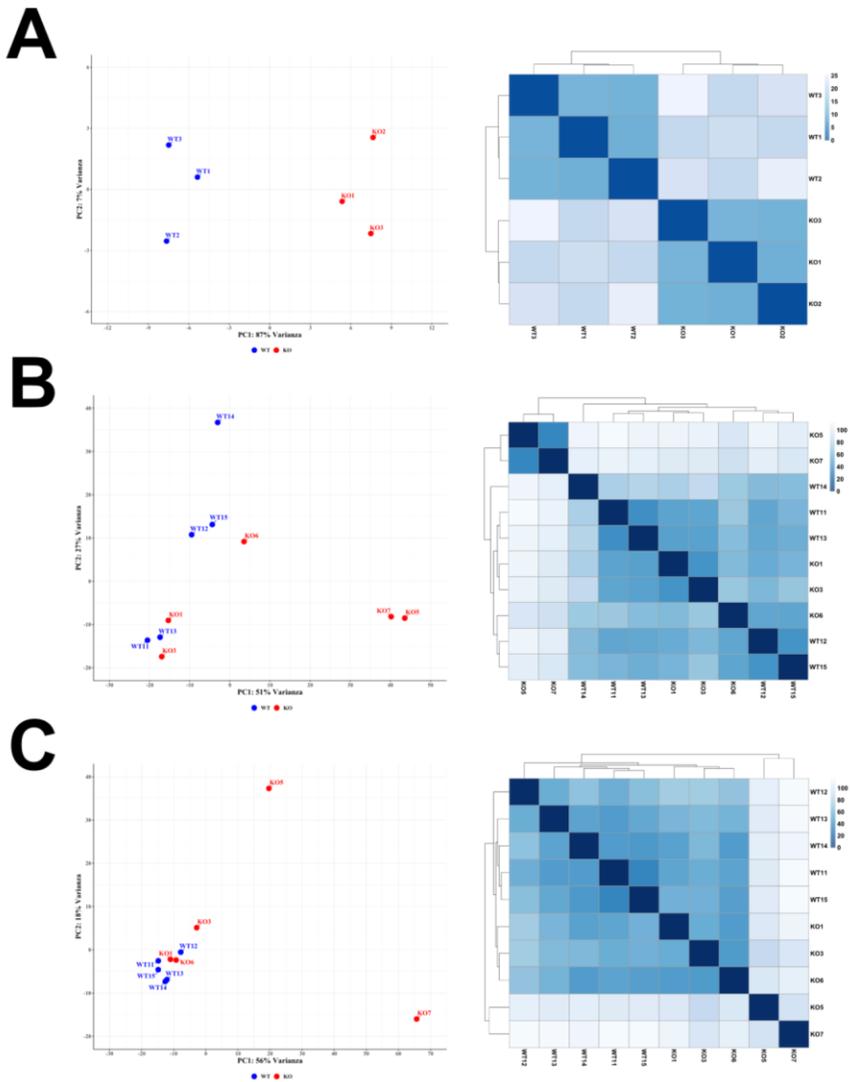
## FASTQC: Niveles de duplicación de secuencia



**Figura 31.** Análisis de duplicación de lecturas de los ficheros FASTQC derivados de secuenciación ARN-Seq realizado por la herramienta FASTQC en las muestras de MEF.

### 5.2.5 Selección de las muestras para el análisis de expresión diferencial

Antes de comenzar el análisis de expresión diferencial de las muestras de MEF, y de tejidos de hígado y de yeyuno de *Mus musculus*, se realizó un análisis de componentes principales (PCA) y un *clustering* jerárquico con el fin de explorar la variabilidad y la agrupación de las réplicas secuenciadas. En el caso de las muestras de MEF, los WT y los DKO se agrupaban de manera separada en ambos controles de calidad [Figura 32A]. Para los tejidos de ratones, los réplicas DKO separadas considerablemente del resto de controles eran la 5 y la 7, tanto para las muestras de hígado [Figura 32B] como para las muestras de yeyuno [Figura 32C]. Por el contrario, las muestras de los DKO 1, 3 y 6 se distribuían cerca del grupo de los controles.



**Figura 32.** Control de calidad de las muestras de ARN-Seq de MEF (A), de hígado (B) y de intestino (C) de ratones WT DKO *Arid1a/Arid1b*. A la izquierda, se muestra el análisis de componentes principales (PCA) indicándose el porcentaje de varianza para cada componente. A la derecha, se muestra un *clustering* jerárquico de las muestras basado en la distancia a la que se encuentran según su similitud.

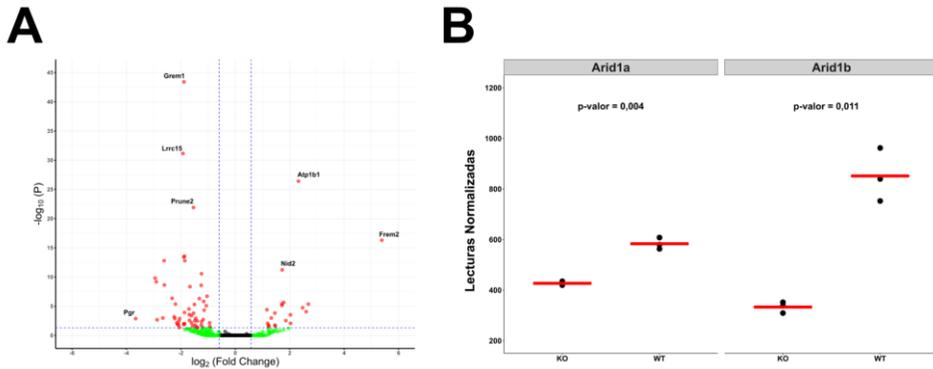
La separación no adecuada de las muestras de los tejidos en cuanto al tratamiento llevó a explorar la presencia de lecturas en los exones de *Arid1a* y *Arid1b* en el visualizador genómico IGV para comprobar si la delección se había realizado satisfactoriamente. Así, se pudo observar que no existían lecturas alineadas en los exones delecionados para los DKO 5 y 7, pero sí para el resto de KO al igual que los controles [Figura S3]. Por ello, con el fin de comparar de

manera robusta el perfil transcriptómico de las muestras KO y las muestras controles, se decidió no incorporar en el análisis de expresión diferencial las muestras DKO 1, 3 y 6. Las muestras finalmente seleccionadas para los análisis posteriores fueron todos los controles y los DKO 5 y 7.

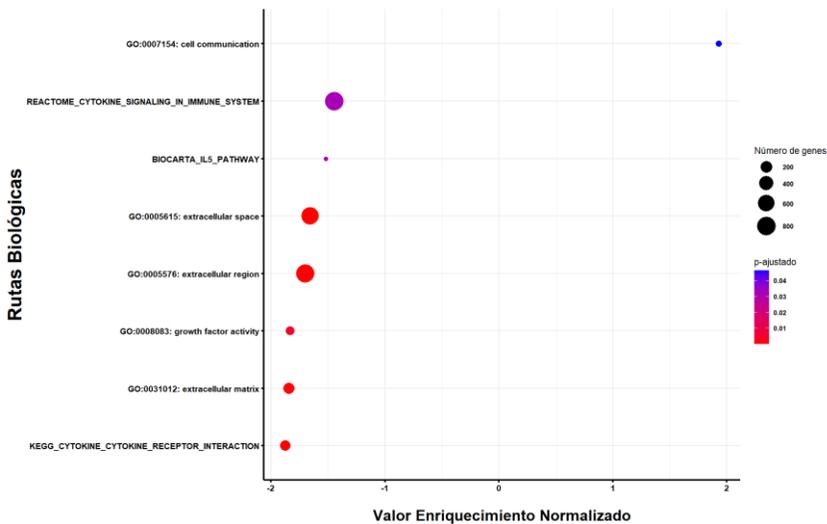
### 5.2.6 Análisis de expresión diferencial y GSEA de MEF

En el análisis de expresión diferencial realizado con *DESeq2*, se tuvo en cuenta la covariable asociada a la réplica debido a que las muestras no se agrupaban apropiadamente sin eliminar este efecto *batch*. Estableciendo como umbrales un valor absoluto de  $FC > 1,5$  y de  $FDR < 0,05$ , se identificaron 74 genes diferencialmente expresados de manera significativa en los MEF DKO *Arid1a/Arid1b* en comparación con los WT, de los cuales 20 se encontraban sobre-expresados y 54 infraexpresados [Figura 33A]. El gen que mostró una mayor sobre-expresión fue la proteína de matriz extracelular *Frem2* ( $FC = 41,79$ ,  $FDR = 5,15 \times 10^{-17}$ ), mientras que el que tuvo una menor expresión en el modelo DKO fue el receptor de progesterona *Pgr* ( $FC = 12,66$ ,  $FDR = 1,33 \times 10^{-3}$ ). En cuanto al grado de significancia, el gen que mostró un cambio más significativo fue la citoquina *Grem1*, cuya expresión se encontraba regulada a la baja en el modelo DKO ( $FC = 3,70$ ,  $FDR = 1,33 \times 10^{-44}$ ). Al realizar una comparación de las lecturas normalizadas de los genes delecionados en el modelo DKO con los ratones WT, tanto la cuantificación de *Arid1a* (p-valor = 0,004) como de *Arid1b* (p-valor = 0,011) fueron significativamente menores [Figura 33B].

Una vez realizado el análisis de expresión diferencial de las muestras de MEF, se llevó a cabo un análisis de enriquecimiento biológico para conocer, en base a los valores umbrales establecidos en la anterior sección, las rutas que se encontraban desreguladas de manera significativa en el modelo DKO *Arid1a/Arid1b*. Dada la baja cantidad de genes expresados diferencialmente de manera significativa obtenidos, el número de rutas biológicas enriquecidas fue escaso, destacando la presencia de aquellas relacionadas con la señalización de citoquinas ( $VEN = -1,45$ ,  $FDR = 3,18 \times 10^{-2}$ ) e interacción de sus receptores ( $VEN = -1,87$ ,  $FDR = 2,88 \times 10^{-5}$ ), que se encontraban reguladas a la baja. La única ruta enriquecida positivamente fue el término GO:0007154 relacionado con la comunicación celular ( $VEN = 1,93$ ,  $FDR = 4,62 \times 10^{-2}$ ), que contenía los genes implicados en señalización *Frem2*, *Jag1* y *Enpp5* [Figura 34].



**Figura 33.** (A) *Volcano plot* de genes expresados diferencialmente entre los MEF DKO *Arid1a/Arid1b* y los WT. La línea horizontal punteada azul representa un umbral FDR = 0,05, mientras que las líneas verticales representan los umbrales para el FC = -1,5/+1,5. Los puntos rojos representan aquellos genes que superan ambos valores umbrales establecidos, mientras que los puntos verdes solo el umbral de FC. (B) Comparación del número de lecturas normalizadas de *Arid1a* y *Arid1b* entre las muestras KO y WT. Se muestra el p-valor derivado del prueba t de Student.

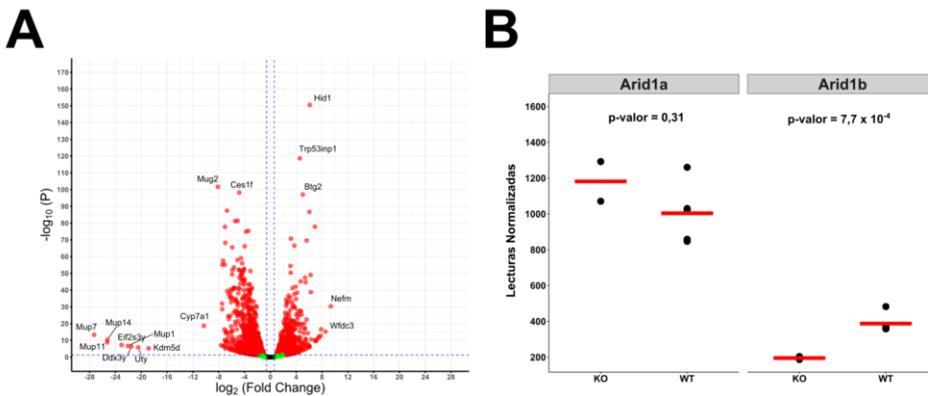


**Figura 34.** Análisis de enriquecimiento biológico de los MEF DKO *Arid1a/Arid1b* en comparación con los WT. Se muestran las rutas biológicas comunes colapsadas. El tamaño de los puntos indica el número de los genes, mientras que el color el p-valor ajustado asociado.

### 5.2.7 Análisis de expresión diferencial y GSEA de los tejidos de hígado y yeyuno

Comparando los tejidos KO5 y KO7 de *Arid1a/Arid1b* de hígado con el modelo WT, según los umbrales establecidos ( $|FC| > 1,5$ ;  $FDR < 0.05$ ), se obtu-

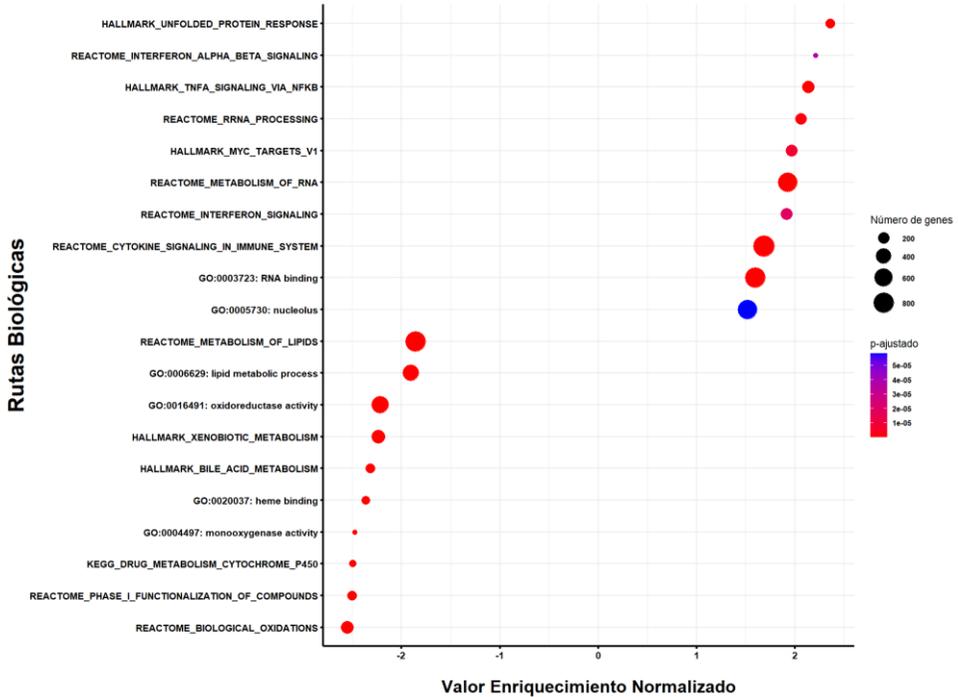
vieron 719 genes sobre-expresados y 1274 regulados a la baja de manera significativa [Figura 35A]. Entre los genes más regulados a la baja, cabe destacar la presencia de las proteínas urinarias mayores *Mup7* ( $\log_2FC = -27,31$ ,  $FDR = 4,56 \times 10^{-14}$ ), *Mup11* ( $\log_2FC = -25,29$ ,  $FDR = 1,48 \times 10^{-9}$ ), *Mup14* ( $\log_2FC = -27,25$ ,  $FDR = 3,51 \times 10^{-11}$ ) y *Mup1* ( $\log_2FC = -22,08$ ,  $FDR = 9,92 \times 10^{-9}$ ). Por otro lado, los tres genes más sobre-expresados fueron el polipéptido de neurofilamentos *Nefm* ( $\log_2FC = 9,37$ ,  $FDR = 4,56 \times 10^{-14}$ ), el dominio 10 de la proteína WAP *Wfdc3* ( $\log_2FC = 8,5$ ,  $FDR = 6,32 \times 10^{-16}$ ) y el ligando anfiregulina para EGFR *Areg* ( $\log_2FC = 7,86$ ,  $FDR = 2,26 \times 10^{-17}$ ). Entre los genes que mostraron un mayor grado de significancia en el análisis de expresión diferencial, se encuentran la proteína *Hid1* ( $\log_2FC = 6,11$ ,  $FDR = 3,01 \times 10^{-151}$ ) y el factor proapoptótico *Trp53inp1* ( $\log_2FC = 4,56$ ,  $FDR = 2,13 \times 10^{-119}$ ). En los hígados DKO, las lecturas normalizadas de los genes deletionados en cuestión fueron menores significativamente únicamente para *Arid1b* (p-valor =  $7,7 \times 10^{-4}$ ). [Figura 35B].



**Figura 35.** (A) *Volcano plot* de genes expresados diferencialmente entre los tejidos de hígado de ratón DKO *Arid1a/Arid1b* y de los WT. La línea horizontal punteada azul representa un umbral  $FDR = 0,05$ , mientras que las líneas verticales representan los umbrales para el  $FC = -1,5/+1,5$ . Los puntos rojos representan aquellos genes que superan ambos valores umbrales establecidos, mientras que los puntos verdes solo el umbral de  $FC$ . (B) Comparación del número de lecturas normalizadas de *Arid1a* y *Arid1b* entre las muestras KO y WT. Se muestra el p-valor derivado de la prueba t de Student.

El análisis de enriquecimiento biológico mostró rutas reguladas a la baja en el modelo KO relacionadas con funciones hepáticas, tales como el metabolismo de los lípidos ( $VEN = -1,86$ ,  $FDR = 4,63 \times 10^{-10}$ ), el metabolismo de los ácidos biliares ( $VEN = -2,32$ ,  $FDR = 7,61 \times 10^{-8}$ ) o el metabolismo de los xenobióticos ( $VEN = -2,23$ ,  $FDR = 1,27 \times 10^{-10}$ ). En el caso de las rutas enriquecidas positivamente, se puede observar la presencia de rutas de señalización relacionadas con el interferón ( $VEN = 1,91$ ,  $FDR = 1,87 \times 10^{-5}$ ), la respuesta a proteínas

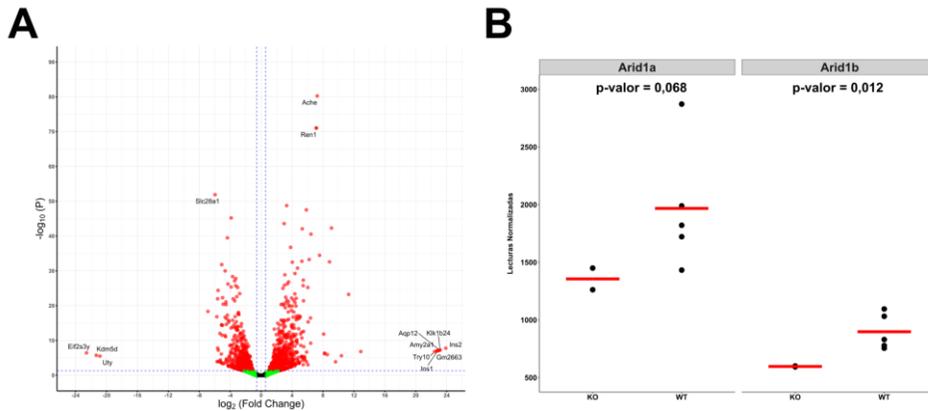
mal plegadas (VEN = 2,36, FDR =  $8,11 \times 10^{-9}$ ) y las dianas de *Myc* (VEN = 1,97, FDR =  $6,92 \times 10^{-6}$ ) [Figura 36]. Adicionalmente, cabe destacar una ruta no presente en el gráfico de enriquecimiento relacionada con regulación al alza de la apoptosis (VEN = 1,75, FDR =  $1,66 \times 10^{-3}$ ).



**Figura 36.** Análisis de enriquecimiento biológico de los tejidos de hígado de ratón DKO *Arid1a/Arid1b* en comparación con las WT. Se muestran las 10 rutas biológicas más enriquecidas y menos enriquecidas ordenadas por el grado de significancia tras ser colapsadas. El tamaño de los puntos indica el número de los genes, mientras que el color el p-valor ajustado asociado.

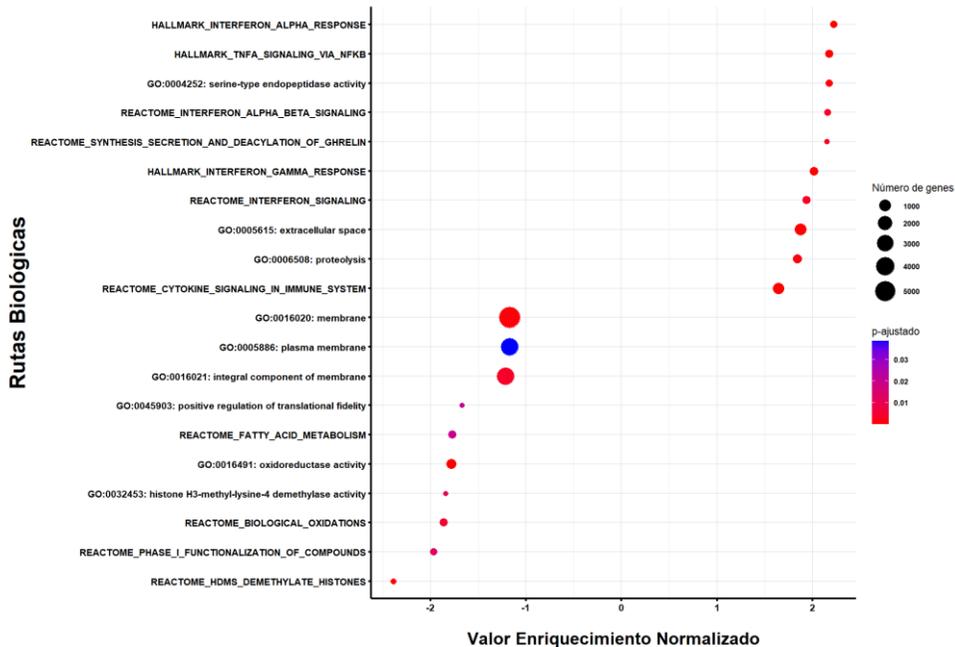
Por último, en cuanto al análisis de expresión diferencial con las muestras DKO y las muestras WT de yeyuno, utilizando los mismos umbrales de los dos análisis anteriores, se obtuvieron 736 genes sobre-expresados y 606 genes regulados a la baja significativos en los ratones KO5 y KO7 [Figura 37A]. Entre los genes con mayor FC, se encuentran la insulina *Ins1* ( $\log_2FC = 23,88$ , FDR =  $1,64 \times 10^{-8}$ ) y *Ins2* ( $\log_2FC = 23,88$ , FDR =  $1,64 \times 10^{-8}$ ), o la proteasa serina calicreína *Klk1b24* ( $\log_2FC = 23,15$ , FDR =  $5,23 \times 10^{-8}$ ). Por otro lado, entre los genes más infraexpresados significativos, se encuentran una subunidad del factor de iniciación de la transcripción *Eif2s3y* ( $\log_2FC = -22,59$ , FDR =  $3,63 \times 10^{-7}$ ), y las demetilinas de histonas *Kdm5d* ( $\log_2FC = -21,32$ , FDR =  $1,81 \times 10^{-6}$ ) y *Uty* ( $\log_2FC = -20,84$ , FDR =  $3,10 \times 10^{-6}$ ). Entre los genes con mayor grado de significancia, cabe destacar la sobre-expresión de la endopeptidasa *Ren1* ( $\log_2FC =$

7,13,  $FDR = 9,28 \times 10^{-72}$ ) y una subunidad del receptor de la acetilcolina *Ache* ( $\log_2FC = 7,25$ ,  $FDR = 5,63 \times 10^{-81}$ ). Para las muestras de yeyuno, de nuevo únicamente la expresión normalizada de *Arid1b* fue significativamente menor en el modelo DKO (p-valor = 0,012) [Figura 37B].



**Figura 37. (A)** *Volcano plot* de genes expresados diferencialmente entre los tejidos de yeyuno de ratón DKO *Arid1a/Arid1b* y de los WT. La línea horizontal punteada azul representa un umbral  $FDR = 0,05$ , mientras que las líneas verticales representan los umbrales para el  $FC = -1,5/+1,5$ . Los puntos rojos representan aquellos genes que superan ambos valores umbrales establecidos, mientras que los puntos verdes solo el umbral de  $FC$ . **(B)** Comparación del número de lecturas normalizadas de *Arid1a* y *Arid1b* entre las muestras KO y WT. Se muestra el p-valor derivado de la prueba t de Student.

El análisis de enriquecimiento biológico de los tejidos de yeyuno mostró rutas de señalización enriquecidas positivamente, tales como la respuesta a interferón- $\alpha$  ( $VEN = 2,22$ ,  $FDR = 5,59 \times 10^{-6}$ ) o a interferón- $\gamma$  ( $VEN = 2,01$ ,  $FDR = 8,12 \times 10^{-5}$ ), y la secreción de la hormona gástrica grelina ( $VEN = 2,15$ ,  $FDR = 3,34 \times 10^{-3}$ ). En el caso de las rutas biológicas reguladas a la baja, se observó un enriquecimiento de términos GO relacionadas con la membrana plasmática ( $VEN = -1,17$ ,  $FDR = 0,038$ ), el metabolismo de los ácidos grasos ( $VEN = -1,76$ ,  $FDR = 0,019$ ), y actividad demetilasa de histonas ( $VEN = -2,39$ ,  $FDR = 2,31 \times 10^{-4}$ ). Además de las no mostradas en el gráfico de enriquecimiento, cabe destacar el enriquecimiento positivo de rutas relacionadas con respuesta a la inflamación ( $VEN = 1,83$ ,  $FDR = 9,91 \times 10^{-3}$ ) y la ruta de señalización IL6-JAK-STAT3 ( $VEN = 2,03$ ,  $FDR = 0,02$ ) [Figura 38].



**Figura 38.** Análisis de enriquecimiento biológico de los tejidos de yeyuno de ratón DKO *Arid1a/Arid1b* en comparación con las WT. Se muestran las 10 rutas biológicas más enriquecidas y menos enriquecidas ordenadas por el grado de significancia tras ser colapsadas. El tamaño de los puntos indica el número de los genes, mientras que el color el p-valor ajustado asociado.

### 5.3 Discusión

Actualmente, la relación biológica causa-efecto entre la alteración del complejo remodelador de la cromatina SWI/SNF y la aparición de enfermedades, tales como el cáncer, aún se encuentra por dilucidar. Entre las razones que dificultan este análisis, se encuentra la redundancia entre las subunidades parálogas y la amplia variedad de subcomplejos SWI/SNF con actividad residual (Helming et al., 2014; Jones et al., 2022). Para comprender con mayor exactitud si la ausencia de un complejo cBAF funcional conduce al desarrollo tumoral, se decidió estudiar las consecuencias fisiológicas, bioquímicas y transcriptómicas de la pérdida de ARID1A y ARID1B en modelos *in vitro* e *in vivo* de ratón.

El sistema de delección utilizado en la generación de los modelos DKO disminuyó satisfactoriamente las lecturas normalizadas de *Arid1b* en todos los modelos. Por el contrario, la bajada de lecturas normalizadas de *Arid1a* únicamente fue significativa en el modelo *in vitro*. Sin embargo, ninguno de ambos genes estuvo desregulado significativamente, según los umbrales establecidos, en el análisis de expresión diferencial con la herramienta DESeq2. La leve disminución de los niveles de ARNm de ambos genes podría deberse a la baja eficiencia del mecanismo de degradación mediado por mutaciones terminadoras (NMD, del inglés, “*Nonsense Mediated Decay*”). Este sistema permite eliminar ARNm con codones de terminación prematuros con el fin de evitar la traducción a proteínas truncadas con posibles efectos tóxicos para la célula (Brognia & Wen, 2009). Lindeboom et al demostraron que la eficiencia del mecanismo NMD puede variar dependiendo de la localización de la mutación terminadora en referencia al complejo de empalme de exones, la distancia al codón de iniciación o la presencia cercana de motivos específicos de unión de proteínas específicas. Adicionalmente, obtuvieron un subgrupo de genes supresores tumorales, que se encuentran alterados frecuentemente por mutaciones terminadoras, cuya secuencia predijeron ser insensible al mecanismo NMD. Entre estos genes se encuentra *ARID1A*, indicando que sus niveles de ARNm son degradados de manera ineficiente. Consecuentemente, las proteínas generadas a partir de estos ARNm alterados por las mutaciones terminadoras pueden encontrarse inactivas, ser ineficientes en cuanto a su función biológica o actuar como dominantes negativos (Lindeboom et al., 2016). En nuestro estudio, la presencia de lecturas en el resto de los exones no delecionados de *Arid1a* como *Arid1b* podría deberse a un mecanismo NMD ineficiente de degradación de ARNm. Aunque en el estudio de Lindeboom no se mostraba el mismo resultado para *ARID1B*, la similitud de secuencias respecto a *ARID1A* detallada en el Capítulo I podría justificar su elusión del mecanismo NMD. Aunque no se ha demostrado en este Capítulo V, las

proteínas generadas a partir de los ARNm alterados de *Arid1a* y *Arid1b* podrían perder su función biológica provocando la desestabilización del complejo cBAF en los modelos estudiados.

En cuanto a los resultados histológicos de los hígados e intestinos presentados en este Capítulo V, se debe tener en cuenta que muestran una tendencia de los signos patológicos observados en los modelos de ratones DKO *Arid1a/Arid1b*, pero no en todas las réplicas biológicas. Es importante considerar que la penetrancia del genotipo puede ser baja en la obtención del fenotipo patológico tras la delección. Adicionalmente, la eficiencia del sistema de eliminación podría ser variable en las distintas réplicas biológicas. Por ejemplo, si no todas las células madre que forman parte de la cripta intestinal pierden *Arid1a* y *Arid1b*, aquellas que mantuvieran ambos genes sin afectar podrían repoblar estas estructuras de nuevas células madre WT para reconstituir las vellosidades intestinales.

Inicialmente, para analizar las rutas biológicas desreguladas obtenidas tras la desestabilización del complejo cBAF mediante la delección de los genes *Arid1a* y *Arid1b*, se realizó una secuenciación ARN-Seq de los MEF WT y DKO *Arid1a/Arid1b*. En el momento del estudio, se eligió este modelo *in vitro* disponible en el laboratorio sin aún conocer los tejidos mayormente afectados en el modelo *in vivo*. En este caso, los resultados del ARN-Seq de los MEF no fueron concluyentes dado que todas las muestras secuenciadas poseían un número alto de lecturas asociadas a ARNr. En consecuencia, se obtuvieron un menor número de genes expresados diferencialmente y un menor número de rutas biológicas enriquecidas de manera significativa en comparación con el ARN-Seq de los tejidos. Esto se atribuyó a un posible error del protocolo de secuenciación en el proceso de depleción de ARNr. Las lecturas asociadas a ARNr podrían estar enmascarando la expresión de PCG, sobre todo, aquellos que se expresan en niveles muy bajos.

Una vez analizados microscópicamente los tejidos del modelo *in vivo* DKO *Arid1a/Arid1b*, se decidió realizar una secuenciación ARN-Seq de tejidos de hígado y yeyuno de los ratones más afectados histológicamente para explorar las rutas biológicas desreguladas. Una limitación del estudio fue la alta similitud que presentaron las muestras KO1, KO3 y KO6 con las muestras WT en ambos tejidos, posiblemente debido a una baja eficiencia del sistema de delección de *Arid1a* y *Arid1b*. Así, los resultados de expresión diferencial se deben a las muestras KO5 y KO7, que presentaron un mayor número de genes expresados diferencialmente en hígado en comparación con las muestras de yeyuno.

## CAPÍTULO V: Implicación de ARID1A y ARID1B en cáncer

Wang et al. demostraron que la delección de *Arid1a* y *Arid1b* en el hígado provoca un incremento de la proliferación y disfunción hepática, y acelera el desarrollo tumoral en los ratones DKO (Z. Wang et al., 2020). Aunque utilizan el mismo sistema de delección usado en este Capítulo V, los resultados de supervivencia que obtienen en los ratones DKO (N = 15) son menos drásticos que los nuestros (N = 54). Sus ratones DKO comienzan a fallecer a partir del día 10 de manera similar a nuestro estudio, pero los que sobreviven más de un mes llegan a alcanzar una edad adulta de más de 6 meses. En nuestro caso, con una población mayor analizada, todos los ratones DKO fallecen antes de los dos primeros meses tras el tratamiento con tamoxifeno.

En relación con los genes expresados diferencialmente en los tejidos de hígado de los ratones KO5 y KO7 *Arid1a/Arid1b* se observa una familia de proteínas regulada a la baja denominada MUP (del inglés, “*Major Urine Protein*”). Las proteínas MUP, producidas mayoritariamente por el hígado, se encuentran implicadas en la estabilización de compuestos volátiles, tales como las feromonas, y en la regulación del metabolismo de los nutrientes (Tirindelli et al., 2009; Zhou & Rui, 2010). En cuanto a la última función mencionada, se ha demostrado que la restricción calórica en ratones reduce notablemente la expresión de *Mup1* en el hígado (Hui et al., 2009). En contraposición a los niveles de glucosa observados en nuestros ratones DKO, se ha observado que la expresión de *Mup1* inhibe los genes relacionados con la gluconeogénesis y disminuye la producción de glucosa en hepatocitos (Zhou et al., 2009). Por tanto, cabría esperar que la regulación a la baja de los genes de la familia MUP podría reactivar la producción de glucosa en el hígado. Sin embargo, debido a la posible ineficiente absorción de nutrientes llevado a cabo en el intestino delgado afectado y los bajos niveles de glucógeno en el hígado observados en estos ratones, los niveles de glucosa en sangre no se pueden incrementar en el modelo DKO aunque los genes *Mup* se encuentren sobre-expresados. Por otro lado, la disfunción hepática de los ratones DKO también se encuentra justificada por un enriquecimiento negativo de rutas biológicas relacionadas el metabolismo de la bilis y con los citocromos P450. Los genes que codifican para los polipéptidos de las distintas familias que componen hemoproteínas citocromos P550 se encuentran regulados a la baja según el análisis de expresión diferencial realizado. Los citocromos P450 son enzimas monooxigenasas que oxidan metabolitos tanto endógenos como exógenos, entre ellos, los xenobióticos (Hart et al., 2009). Se ha observado que la delección de *Arid1a* en hígado de ratones reduce los niveles de ARNm de genes que componen la familia de citocromos P450 (Sun et al., 2016, 2017). En consecuencia, los términos biológicos relacionados con el metabolismo de xenobióticos, las reac-

ciones enzimáticas de oxidación y la unión a hemoproteínas se encuentran sub-representados en el análisis de enriquecimiento para los ratones DKO.

A diferencia del estudio de Wang et al., no se observó fenotípicamente ningún crecimiento tumoral en los hígados de los ratones DKO. Sin embargo, de manera similar, nuestro análisis de enriquecimiento mostró rutas reguladas positivamente relacionadas con la apoptosis y otras asociadas al desarrollo tumoral, por ejemplo, las dianas de *Myc*. En este contexto, cabe destacar la sobre-expresión de *Cdkn1a* ( $\log_2FC = 3,53$ ,  $FDR = 2,15 \times 10^{-19}$ ) significativa en las muestras DKO en comparación con las WT. Por último, en relación con los signos de esteatosis observados en las imágenes de microscopía y los niveles altos de colesterol, se obtuvieron rutas relacionadas reguladas a la baja significativamente relacionadas con el metabolismo de los ácidos grasos.

Por otra parte, Wang et al. no analizaron los posibles efectos de la doble delección de *Arid1a* y *Arid1b* en intestino delgado (Z. Wang et al., 2020). Aunque se desconoce el rol que puede ejercer *Arid1a* en la homeostasis del epitelio intestinal, se ha publicado en los últimos años que su delección puede inducir cáncer colorrectal (Kamori et al., 2022; Zhao et al., 2022). Dado que el complejo remodelador de la cromatina cBAF controla programas transcripcionales de diferentes líneas celulares en distintos estadios, su alteración puede conducir a la pérdida del control regulatorio de los *enhancers* asociados. Así, se ha demostrado que la pérdida de *Arid1a* puede provocar la pérdida de unión del complejo SWI/SNF a miles de *enhancers* activos alterando la expresión de los genes controlados por éstos (Mathur et al., 2017). En 2019, Hiramatsu et al. publicaron un estudio realizado sobre la delección de *Arid1a* específica de intestino en ratones. Como resultado, la pérdida de la expresión de esta subunidad supuso un desbalanceo en el mantenimiento de las células madre en el epitelio intestinal. A nivel histológico, se produjo un acortamiento de las vellosidades intestinales y una inflamación de las criptas en el intestino delgado, cuya intensidad se incrementaba con el envejecimiento del ratón (Hiramatsu et al., 2019). De manera similar, nuestros modelos DKO presentaron un acortamiento y desorganización de las vellosidades, y una proliferación celular disminuida en las criptas de Lieberkühn. Actualmente, se conoce que la subunidad ATPasa SMARCA4 modula la renovación de las células madre en el intestino delgado a través de la ruta de señalización Wnt (M. Liu et al., 2019). Por tanto, la delección de las proteínas *Arid1* en nuestro modelo DKO podría generar un complejo cBAF no funcional o residual que desestabilizaría la homeostasis de las células madre. Además, esto ha sido confirmado en nuestro estudio con la bajada de expresión génica de los marcadores de células madre intestinales *Lgr5* y *Ascl2* mediante qPCR en muestras de criptas purificadas. Ambos genes no fueron estadísticamente significativos en el análisis de ex-

presión diferencial de los datos procedentes del ARN-Seq debido posiblemente al uso de muestras enteras de intestino delgado. En cambio, sí se pudo confirmar una disminución de la expresión significativa de otro marcador de células madre intestinales denominado *Olfm4* ( $\log_2FC = -5,56$ ,  $FDR = 4,09 \times 10^{-25}$ ).

En cuanto al análisis de enriquecimiento biológico realizado en las muestras de intestino, se observaron rutas enriquecidas positivamente relacionadas con la señalización inmunológica mediada por interferón. Estas citoquinas se encuentran implicadas en el reconocimiento de antígenos ante microorganismos desconocidos para el organismo, en la proliferación y diferenciación celular, y en el mantenimiento de la función de la barrera intestinal (Katlinskaya et al., 2016). El mantenimiento de la pared intestinal resulta crucial para una absorción eficiente de los nutrientes (Schoultz & Keita, 2020), por lo que su alteración podría estar influyendo de manera indirecta en la pérdida de peso del ratón observada y la disminución de niveles de glucosa. Además, se obtuvo una ruta relacionada con respuesta a la inflamación regulada al alza, y la ruta IL6-JAK-STAT3, implicada en la regulación de la permeabilidad de las células epiteliales (L. Pang et al., 2021) En el caso de rutas reguladas a la baja, destacó la presencia de términos GO relacionados con la membrana plasmática. En conjunto, estas rutas biológicas desreguladas en los modelos DKO sugieren una activación de vías inmunológicas y la alteración de la barrera epitelial intestinal.

En general, los resultados obtenidos para el modelo *in vivo*, tanto histológica como transcriptómicamente, indican que la delección de *Arid1a* y *Arid1b* producen una disfunción en el hígado y en la barrera intestinal. La eliminación de ambos genes produjo una muerte drástica del ratón sin aparición de tumores. En el futuro, se pretende analizar los promotores y *enhancers* regulados por *Arid1a* y *Arid1b* que son relevantes para el control de la expresión génica de los tejidos afectados.

# Conclusiones

# Conclusiones

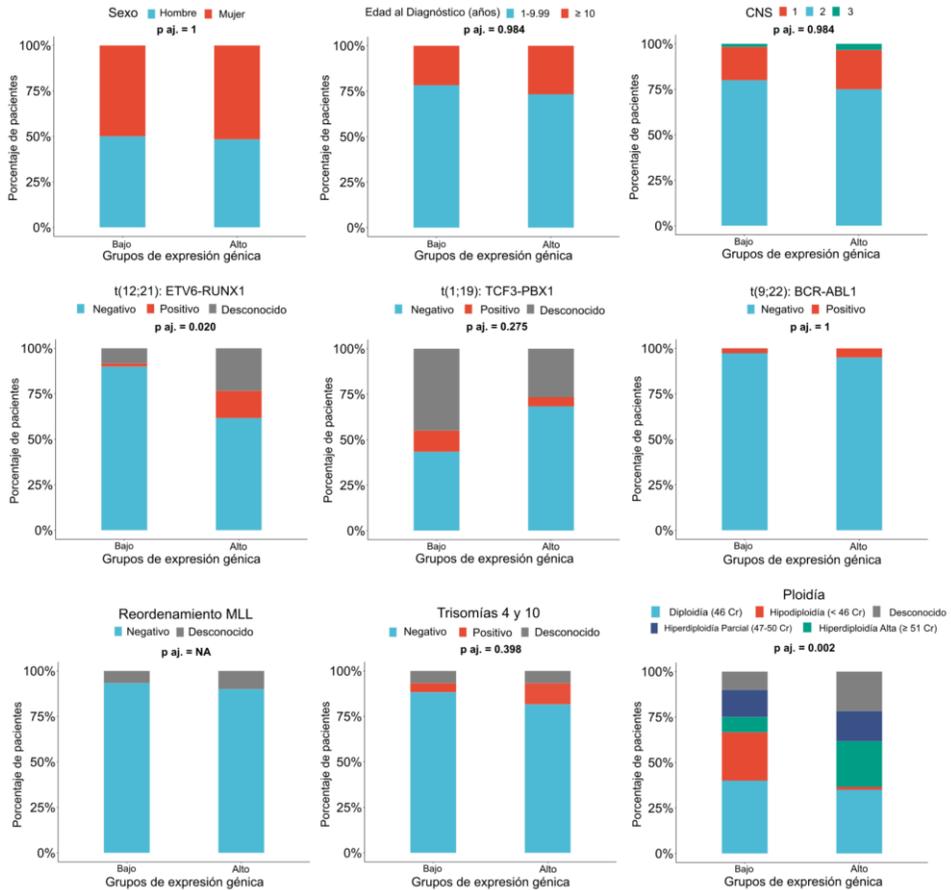
1. El ARNinc *AL133346.1* y su gen solapante *CCN2* se sobre-expresan significativamente en comparación con individuos sanos y pacientes pediátricos con leucemia linfoblástica aguda de células T, por lo que pueden considerarse como biomarcadores diagnósticos de la leucemia linfoblástica aguda de células B pediátrica.
2. La correlación de expresión significativa entre *AL133346.1* y *CCN2* en leucemia linfoblástica aguda de células B pediátrica podría deberse a una regulación en *cis* producida por el ARNinc o por el efecto de elementos reguladores comunes.
3. La sobre-expresión de *CCN2* puede considerarse un biomarcador pronóstico independiente que mejora la supervivencia global de los pacientes pediátricos con leucemia linfoblástica aguda de células B.
4. La expresión de la subunidad ARID2 del complejo SWI/SNF tuvo un impacto significativo en la supervivencia de pacientes con leucemia linfoblástica aguda de células B pediátrica, asociándose mayores niveles con tiempos de supervivencia global más largos.
5. Aunque *ARID2* es conocido como supresor tumoral en otros tipos de cáncer, su mayor sobre-expresión en pacientes con leucemia linfoblástica aguda de células B pediátricos en comparación con muestras sanas reveló su posible papel en el mantenimiento tumoral de esta neoplasia hematológica.
6. La inhibición de la expresión génica de *ARID2* en líneas celulares de leucemia linfoblástica aguda de células B pediátrica y adulta provocó una disminución de la viabilidad celular.
7. La delección de los genes *Arid1a* y *Arid1b* produce una disminución drástica de la supervivencia de los ratones doble *knock-out* en comparación con los silvestres.
8. Los ratones doble *knock-out* de *Arid1a* y *Arid1b* presentaron signos patológicos en el hígado, una alteración estructural de las vellosidades intestinales y una reducción de proliferación de las células madre en las crip-

tas intestinales, señalando la importancia del complejo cBAF en la regulación de la expresión génica en estos tejidos.

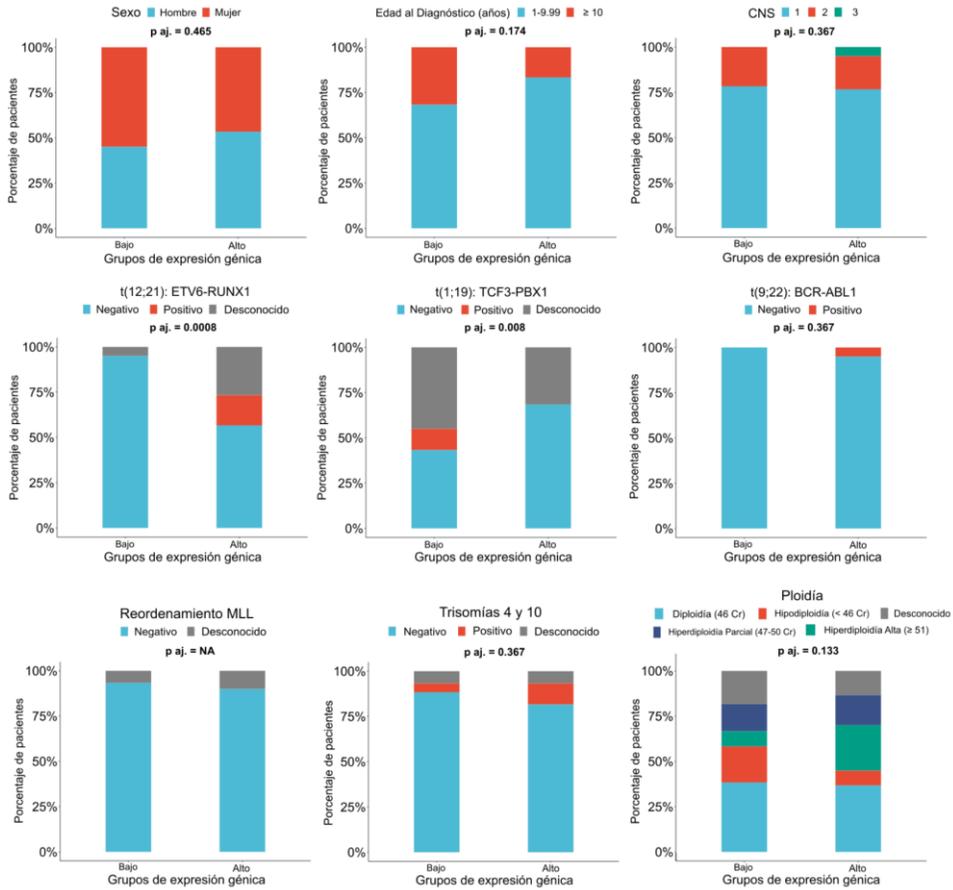
9. Entre las rutas celulares biológicas desreguladas por la delección de *Arid1a* y *Arid1b* en modelo de ratón, se encuentran la apoptosis, la señalización inmunológica e inflamatoria, y alteraciones en el metabolismo de los nutrientes.

# Material Suplementario

# Material Suplementario



**Figura S1.** Comparación de características clínicas según la expresión de *AL13346.1* en pacientes pediátricos con LLA-B procedentes del conjunto de datos TARGET. Se muestran los p-valores ajustados por FDR de las pruebas exactas de Fisher o de las pruebas chi-cuadrado. Abreviaturas: Cr: Cromosoma; CNS: Sistema Nervioso Central; CNS 1 significa ausencia de afectación del SNC, CNS 2 es un nivel muy bajo y CNS 3 indica afectación definitiva del SNC NA: no disponible; MLL: gen MLL; t(12;21): translocación (12;21); t(1;19): translocación (1;19); t(9;22): translocación (9;22).

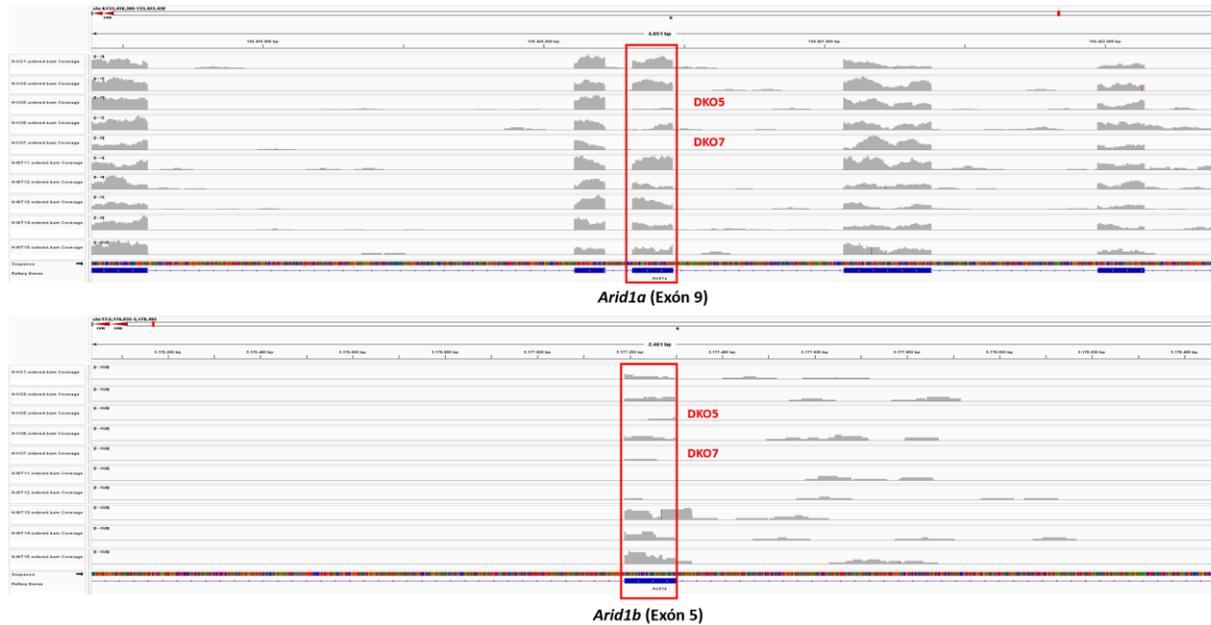


**Figura S2.** Comparación de características clínicas según la expresión de *CCN2* en pacientes pediátricos con LLA-B procedentes del conjunto de datos TARGET. Se muestran los p-valor ajustados por FDR de las pruebas exactas de Fisher o de las pruebas chi-cuadrado. Abreviaturas: Cr: Cromosoma; CNS: Sistema Nervioso Central (SNC); CNS 1 significa ausencia de afectación del SNC; CNS 2 es un nivel muy bajo y CNS 3 indica afectación definitiva del SNC; NA: no disponible; MLL: gen MLL; t(12;21): translocación (12;21); t(1;19): translocación (1;19); t(9;22): translocación (9;22).

**Figura S3.** Capturas de alineamiento de lecturas en los exones 9 de *Arid1a* y 5 de *Arid1b* delecionados procedente de la secuenciación de ARN-Seq de tejidos de hígado (A) y yeyuno (B) de ratón DKO *Arid1a/Arid1b* y WT. De arriba hacia abajo, se muestra KO1, KO3, KO5, KO6, KO7, WT11, WT12, WT13, WT14, WT15.

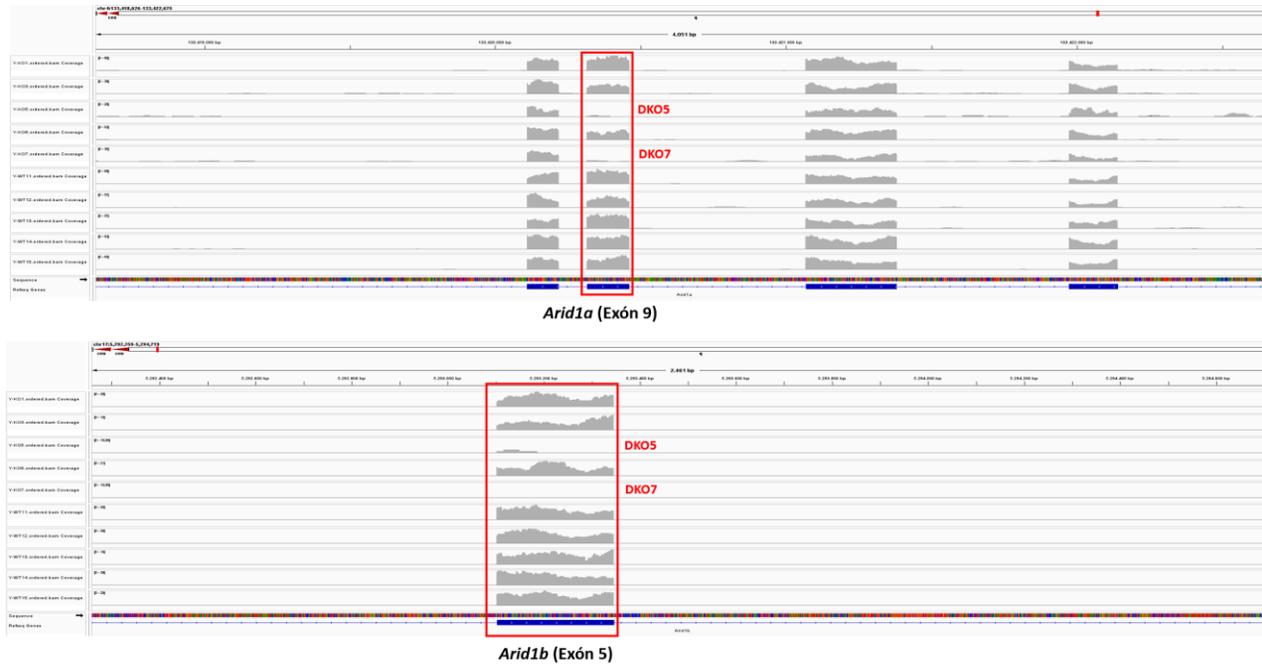
**A**

### Archivos BAM - Hígado



**B**

## Archivos BAM - Intestino delgado (yeyuno)



**Tabla S1.** 48 ARNlnc, representados por 50 sondas, que presentaron una expresión diferencial significativa entre las muestras de pacientes pediátricos con LLA-B y las muestras de médula ósea sanas ( $|FC| > 1,5$ ,  $FDR < 0,05$ ). Los ARNlnc sobre-expresados se muestran en rojo y los ARNlnc regulados a la baja se muestran en verde, según el valor FC (del inglés, “Fold Change”). FDR: “False Discovery Rate”.

Nombre del Gen	FC	p-valor	FDR	Identificador	Fuente	Secuencia Sonda
XLOC_007191	7,011	3,31E-04	1,25E-02	TCONS_00015148	Cabili et al	AGCTGCCTGCTTAGAAGGACTTTGGAAGCAAGACAGAGCAGTGTCT
PCNA-AS1	6,869	6,63E-05	3,17E-03	NR_028370	RefSeq	GAAAGACAACGACCACTCTGCTACGCCTGCAACCGTTTAATGCCGC
AL133346,1	5,155	2,32E-05	1,48E-03	ENST00000435287	gencode	GTAACCTGGACAACCTGAGGCATTAATTGAGGATCAATATGATGAC
HCP5	4,496	4,06E-05	2,30E-03	NR_040662	RefSeq	GCCCTGAGGCAAACATGAATTAAGTGTGTGGATCCCACATGAA
LOC144481	4,151	4,28E-06	4,37E-04	uc001tcu.3	UCSC_knowngene	TATTTTACCACAGGTACCAATCACCTCCAGATCTAAGAATGGCTT
RP11-334C17,5	4,011	5,81E-04	1,78E-02	ENST00000573346	gencode	GGACCACTGCCTGCACGCCTTGGTCAAAAACAAAGCCACACACCAGA
OMG	3,606	1,02E-03	2,84E-02	ENST00000584094	gencode	CAACTGCCTGCATTCTTATCTCAGTTTCTTATTATTGCTGAATGAC
RP1-159A19,4	3,136	5,35E-04	1,73E-02	ENST00000443579	gencode	GGTTTGAATGTACTACTGAGTGGTCTTCATAAATTACCTCTCTGC
RP11-164H13,1	3,009	8,38E-06	7,25E-04	ENST00000553445	gencode	TTCAGCCTGTGACTGCCACCAGAGATGTCTTCTCGCGGTAGCTG
LOC100130476	2,892	1,81E-05	1,21E-03	NR_049793	RefSeq	CTGATGTAAGTCTTGTGCATCTTGCTGCCAGGATTTGTGACAAG
RP11-356I2,4	2,856	5,69E-04	1,78E-02	ENST00000448942	gencode	CTCACTTCATAAGAGCAGTAAACTAAACCAAGGAATGTGAATT
LOC100887755	2,732	8,79E-04	2,54E-02	NR_051976	RefSeq	GGATGCCTGCTACTGTCACTGTTGTCAATTAGATAAATTAATGAAC
PTENP1	2,686	5,15E-05	2,63E-03	NR_023917	RefSeq	TGTGAATGCTTCATGTGCTGCCTGCAAGCTTCTTTTTTCTCATTAA
TRAF3IP2-AS1	2,617	4,58E-04	1,56E-02	NR_034110	RefSeq	TCAAGAAGATCAGAACAGGCTTAAATTTGATGCTATAAAACCAATTA
CTD-2012K14,7	2,301	1,70E-04	7,43E-03	ENST00000562846	gencode	GACCGTTCGGTTGTGAGAATGCTGGCCCAATAACATACTCTTTTTT
MYLK-AS1	2,023	8,91E-05	4,14E-03	NR_038266	RefSeq	GAATGACTTTGTGACCAGCAGCTTAAGCCAAATAAACTTTATAAGG
PHBP11	1,951	9,41E-04	2,67E-02	ENST00000420550	gencode	ACAGAAACCAATTATCTTTGACTGCCGTTCTCGACCACGTAAAGTG
RP11-164H13,1	1,947	1,04E-05	7,73E-04	ENST00000547644	gencode	ACCATACATCAGTCATCTGGCAGCAGCTCGAGAAGCATGTCTCCA
RAB11B-AS1	1,896	1,39E-03	3,61E-02	NR_038237	RefSeq	TCATTTTTAAACAAACTGAGGGTAACTTACTAGATCTGTGAACT
RP5-1061H20,3	1,718	1,67E-03	4,05E-02	ENST00000434404	gencode	AGAGCGGCAGGTGGGGCGTTGGGGGCTGAGTCCCGATTTCCCTGA

RPL13AP20	1,574	4,22E-04	1,51E-02	NR_003932	RefSeq	AGAAGAAAATTGACAAATACACAGAAGTCTCAAGACCCACGGACT
RP11-807H22,6	-13,874	1,51E-04	6,79E-03	ENST00000539482	gencode	CCCTGGGGTCTTCCGACAACCTATCCTAATAGACAAATCCACATG
CXCR2P1	-5,796	0	0	NR_002712	RefSeq	GTACTIONTTGCCAAGAACCAAGGCTGAAATTTCTTGCTAAATGAGA
BC127858	-4,441	1,19E-09	2,60E-07	uc002zti,1	UCSC_knowngene	AAACACCCAAGTGTAGGAACCAGAATAATCAGCCCTAATAGAGGG
LINC00152	-3,901	6,20E-08	1,05E-05	NR_024206	RefSeq	TTGAATAACTGGGAGATGAAACAGGAAGCTCTATGACACACTTGAT
ACTN1-AS1	-3,689	7,00E-11	2,70E-08	ENST00000553944	gencode	TCTAAACTGTGGCTCTAAACCTAACACAGAGGGTAATGGTAGTGCT
RP11-124N14,4	-3,521	3,96E-06	4,33E-04	ENST00000437232	gencode	TGTTTCGAAGTCGATGCAAAAAATGAGTTGCTTATTTTCAGTCTCTC
RP11-51B10,4	-3,345	4,45E-04	1,55E-02	ENST00000451920	gencode	TGCAGGGACCTCTTATCATCCATGGCAAGGATGTCCTTCCCCAGA
AC092214,10	-2,775	1,30E-10	4,00E-08	ENST00000421648	gencode	AGTGTGTATTACAGATACAAGCCTTATCTCAATACCTTCTTACCC
ACTR3P2	-2,708	2,14E-04	8,85E-03	ENST00000439585	gencode	TGAGTGGCAGTAGATTGAAGCCAAAATCTATTGATGTACAAGTCGT
RP11-366L20,4	-2,547	2,20E-10	5,60E-08	ENST00000536217	gencode	GGTTTCTGTCTCTTTTGCTTCTCGAAAATACAGACCTACAAAGC
XLOC_000218	-2,483	1,00E-11	5,00E-09	TCONS_00000959	Cabili et al	CATCTTAGTGTGACTAAAAGAGGGAGATCCTAAAGAATTTCAAGGA
GYG1P1	-2,404	4,72E-08	9,03E-06	ENST00000460864	gencode	CAAAAAGTGTCAAAAAGTGATTCCCATGATCCCAACATAACTTATCC
HMGN2P36	-2,307	1,99E-04	8,47E-03	ENST00000529541	gencode	GAGATGCCAAGTGAAGTATGTGCATTTTGTATAACTGTGACTCCT
RP11-393K10,1	-2,237	0	0	ENST00000431627	gencode	GCACAATATCAAGTGTGAGGAAGGAGATAAGTTTTGATTATTCTGC
HMGN2P5	-2,215	1,25E-03	3,36E-02	ENST00000436882	gencode	CCAAAACAGACCAGGCACAGAAAAGCTGAAGGTGCTGGAGATGCCAA
RP5-1042I8,7	-2,078	1,40E-05	9,77E-04	ENST00000566949	gencode	TAGTGTGTTTCTTACCACCTTTGGCTTTGAGCACTCAAAGTGCAGA
RP11-603J24,7	-2,025	8,71E-04	2,54E-02	ENST00000552016	gencode	GTACTIONACTGAAAACCTTGTGACATAAAGAACCAACAAGTGCTC
RP11-293M10,6	-1,946	2,53E-04	1,02E-02	ENST00000560419	gencode	GGCTTCTCTTCCAATGACCACCAATGAGAATGTTATAAAAATATGTC
LINC01133	-1,933	5,52E-05	2,73E-03	NR_038849	RefSeq	GCTTTGTATGTTATTAATGTGTCTCGTCAATGCTGTTGGCATTGT
XLOC_012542	-1,920	1,86E-06	2,19E-04	TCONS_00025724	Cabili et al	GCGGGAGCTGGTGCCTTGGCAGGAGGGACTGGCTTATGCCACCCTG
RP11-315D16,3	-1,812	1,74E-03	4,17E-02	ENST00000566875	gencode	ATGAAGGACAAACCACAGAGAAGATCCACGAGGTTTTCTGCTAAAC
CASPIP1	-1,806	5,42E-04	1,73E-02	ENST00000526345	gencode	TGAAAGGATGACTTTGACAATACGCTTCTACCTTCTCCAGAACAT
RPL23AP7	-1,782	1,06E-05	7,73E-04	NR_024530	RefSeq	TTCCACACTACAGAAAAGATCCCACTTCTTAGAGAACCAGTGA

RP11-403F21,1	-1,643	4,24E-04	1,51E-02	ENST00000443069	gencode	TTCTACATAAACCGAGTCAGCCACCAGATTCAGACCTTTGTGACAA
RP11-211G3,2	-1,628	4,41E-05	2,33E-03	ENST00000450760	gencode	CACATAGAAAACCTTGGAGCCAAAGCATTGGCAAGAGCGGAAAAAA
LINC00853	-1,622	1,76E-07	2,69E-05	NR_047498	RefSeq	GGATCCTTGAAGCACAAATAAACAGATAGTATTTCTGCATGTGTCAA
AC096582,8	-1,621	1,05E-05	7,73E-04	ENST00000413839	gencode	TCTGAGCCAGCTGCACGACAAGTTCGAGCATCTTAAAATGATTCAA
RPL23AP7	-1,602	4,34E-05	2,33E-03	NR_024531	RefSeq	AGACATTTAGGAGAAGAATCCCAGTCTTCTTAGAGAACCCAGTGAC
ARF4P2	-1,514	1,45E-03	3,64E-02	ENST00000430209	gencode	TCTTCAGTCTCTTCCTAACAGGACATGGTATGTTCAAGCCACTTGT

**Tabla S2.** Porcentajes de pacientes LLA-B pediátricos con las distintas covariables clínicas estudiadas, en los subgrupos “Bajo” y “Alto” del ARNlnc *AL133346.1* y el PCG *CCN2*. Se muestran los p-valores ajustados por FDR de las pruebas exactas de Fisher o de las pruebas chi-cuadrado. Abreviaturas: Cr: Cromosoma; CNS: Sistema Nervioso Central (SNC); CNS 1 significa ausencia de afectación del SNC, CNS 2 es un nivel muy bajo y CNS 3 indica afectación definitiva del SNC; NA: no disponible; MLL: gen MLL; t(12;21): translocación (12;21); t(1;19): translocación (1;19); t(9;22): translocación (9;22).

Cohorte TCGA - LLA-B Pediátrica (N = 120)								
	Grupos de Expresión AL133346.1				Grupos de Expresión CCN2			
	Bajo (N = 60)	Alto (N = 60)	p-valor	FDR	Bajo (N = 60)	Alto (N = 60)	p-valor	FDR
<b>Sexo</b>								
Mujer	30 (50,0%)	31 (51,7%)	1	1	33 (55,0%)	28 (46,7%)	0,465	0,465
Hombre	30 (50,0%)	29 (48,3%)			27 (45,0%)	32 (53,3%)		
<b>Edad al Diagnóstico (años)</b>								
1-9.99	47 (78,3%)	44 (73,3%)	0,670	0,984	41 (68,3%)	50 (83,3%)	0,087	0,174
≥ 10	13 (21,7%)	16 (26,7%)			19 (31,7%)	10 (16,7%)		
<b>CNS</b>								
1	48 (80,0%)	45 (75,0%)	0,738	0,984	47 (78,3%)	46 (76,7%)	0,282	0,367
2	11 (18,3%)	13 (21,7%)			13 (21,7%)	11 (18,3%)		
3	1 (1,7%)	2 (3,3%)			0 (0%)	3 (5,0%)		
<b>t(12;21): ETV6-RUNX1</b>								
Negativo	54 (98,2%)	37 (80,4%)	0,005	0,020	57 (100%)	34 (77,3%)	0,0001	0,0008
Positivo	1 (1,8%)	9 (19,6%)			0 (0%)	10 (22,7%)		
<b>t(1;19): TCF3-PBX1</b>								
Negativo	24 (82,8%)	43 (95,6%)	0,103	0,275	26 (78,8%)	41 (100%)	0,002	0,008
Positivo	5 (17,2%)	2 (4,4%)			7 (21,2%)	0 (0%)		
<b>t(9;22): BCR-ABL1</b>								
Negativo	59 (98,3%)	58 (96,7%)	1	1	60 (100%)	57 (95,0%)	0,244	0,367
Positivo	1 (1,7%)	2 (3,3%)			0 (0%)	3 (5,0%)		
<b>Reordenamiento MLL</b>								
Negativo	57 (100%)	53 (100%)	NA	NA	56 (100%)	54 (100%)	NA	NA
Positivo	0 (0%)	0 (0%)			0 (0%)	0 (0%)		
<b>Trisomías 4 y 10</b>								
Negativo	54 (94,7%)	48 (87,3%)	0,199	0,398	53 (94,6%)	49 (87,5%)	0,321	0,367
Positivo	3 (5,3%)	7 (12,7%)			3 (5,4%)	7 (12,5%)		
<b>Ploidía</b>								
Hipodiploidia (< 46 Cr)	16 (29,6%)	1 (2,1%)	0,0002	0,002	12 (24,5%)	5 (9,6%)	0,05	0,133
Diploidia (46 Cr)	24 (44,4%)	21 (44,7%)			23 (46,9%)	22 (42,3%)		
Hiperdiploidia Parcial (47-50 Cr)	9 (16,7%)	10 (21,3%)			9 (18,4%)	10 (19,2%)		
Hiperdiploidia Alta (≥ 51)	5 (9,3%)	15 (31,9%)	5 (10,2%)	15 (28,9%)				

**Tabla S3.** Análisis Cox univariable y multivariable de las covariables clínicas de los pacientes pediátricos con LLA-B derivados del proyecto TARGET de TCGA (N = 120). Los valores NA aparecen debido a que las correspondientes covariables clínicas no fueron incluidas en el análisis Cox multivariable. La covariable relacionada con los reordenamientos MLL no fue incluida debido a que no existían pacientes positivos.

	Análisis Cox Univariable		Análisis Cox Multivariable	
	Cociente de Riesgo (95% IC)	p-valor	Cociente de Riesgo (95% IC)	p-valor
Sexo (Hombre vs Mujer)	1,852 (1,062-3,228)	0,030	1,269 (0,623-2,585)	0,512
Edad al Diagnóstico (≥ 10 vs 1-9,9 años)	1,500 (0,823-2,733)	0,186	1,550 (0,681-3,527)	0,296
CNS (2-3 vs 1)	1,024 (0,538-1,949)	0,943	NA	NA
t(12;21): ETV6-RUNX1 (Positivo vs Negativo)	0,697 (0,250-1,944)	0,490	NA	NA
t(1;19): TCF3-PBX1 (Positivo vs Negativo)	21,840 (6,968-68,480)	$1,22 \times 10^{-7}$	9,401 (2,579-34,274)	$6,85 \times 10^{-4}$
t(9;22): BCR-ABL1 (Positivo vs Negativo)	0,667 (0,092-4,834)	0,689	NA	NA
Trisomías 4 y 10 (Positivo vs Negativo)	0,343 (0,192-1,986)	0,418	NA	NA
Ploidía (Hipodiploidía vs Diploidía)	0,126 (0,030-0,531)	0,005	0,259 (0,033-2,015)	0,197
Ploidía (Hiperdiploidía Parcial vs Diploidía)	0,383 (0,158-0,926)	0,033	0,821 (0,308-2,186)	0,693
Ploidía (Hiperdiploidía Alta vs Diploidía)	0,346 (0,143-0,838)	0,019	0,448 (0,164-1,221)	0,117
Expresión ARID1A (Alto vs Bajo)	1,870 (1,073-3,261)	0,027	1,330 (0,530-3,336)	0,543
Expresión ARID1B (Alto vs Bajo)	0,779 (0,454-1,336)	0,363	NA	NA
Expresión ARID2 (Alto vs Bajo)	0,457 (0,262-0,799)	0,006	0,600 (0,284-1,267)	0,180
Expresión SMARCA4 (Alto vs Bajo)	2,039 (1,168-3,560)	0,012	1,833 (0,753-4,462)	0,182
Expresión PBRM1 (Alto vs Bajo)	0,924 (0,539-1,585)	0,774	NA	NA

# **Bibliografía**

## Bibliografía

- Adcock, I. M., Ford, P., Ito, K., & Barnes, P. J. (2006). Epigenetics and airways disease. *Respiratory Research*, 7(1), 21. <https://doi.org/10.1186/1465-9921-7-21>
- Adnan-Awad, S., Kim, D., Hohtari, H., Javarappa, K. K., Brandstoetter, T., Mayer, I., Potdar, S., Heckman, C. A., Kytölä, S., Porkka, K., Doma, E., Sexl, V., Kankainen, M., & Mustjoki, S. (2021). Characterization of p190-Bcr-Abl chronic myeloid leukemia reveals specific signaling pathways and therapeutic targets. *Leukemia*, 35(7), 1964–1975. <https://doi.org/10.1038/s41375-020-01082-4>
- Andrades, A., Peinado, P., Alvarez-Perez, J. C., Sanjuan-Hidalgo, J., García, D. J., Arenas, A. M., Matia-González, A. M., & Medina, P. P. (2023). SWI/SNF complexes in hematological malignancies: biological implications and therapeutic opportunities. *Molecular Cancer*, 22(1), 39. <https://doi.org/10.1186/s12943-023-01736-8>
- Astori, A., Tingvall-Gustafsson, J., Kuruvilla, J., Coyaud, E., Laurent, E. M. N., Sunnerhagen, M., Åhsberg, J., Ungerback, J., Strid, T., Sigvardsson, M., Raught, B., & Somasundaram, R. (2020). ARID1a Associates with Lymphoid-Restricted Transcription Factors and Has an Essential Role in T Cell Development. *The Journal of Immunology*, 205(5), 1419–1432. <https://doi.org/10.4049/jimmunol.1900959>
- Attaway, M., Chwat-Edelstein, T., & Vuong, B. Q. (2021). Regulatory Non-Coding RNAs Modulate Transcriptional Activation During B Cell Development. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.678084>
- Azad, P., Caldwell, A. B., Ramachandran, S., Spann, N. J., Akbari, A., Villafuerte, F. C., Bermudez, D., Zhao, H., Poulsen, O., Zhou, D., Bafna, V., Subramaniam, S., & Haddad, G. G. (2022). ARID1B, a molecular suppressor of erythropoiesis, is essential for the prevention of Monge's disease. *Experimental & Molecular Medicine*, 54(6), 777–787. <https://doi.org/10.1038/s12276-022-00769-1>
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), 823–837. <https://doi.org/10.1016/j.cell.2007.05.009>

- Bartholomew, B. (2014). Regulating the Chromatin Landscape: Structural and Mechanistic Perspectives. *Annual Review of Biochemistry*, 83(1), 671–696. <https://doi.org/10.1146/annurev-biochem-051810-093157>
- Bateman, N. W., Shoji, Y., Conrads, K. A., Stroop, K. D., Hamilton, C. A., Darcy, K. M., Maxwell, G. L., Risinger, J. I., & Conrads, T. P. (2016). Identification and functional characterization of a novel bipartite nuclear localization sequence in ARID1A. *Biochemical and Biophysical Research Communications*, 469(1), 114–119. <https://doi.org/10.1016/j.bbrc.2015.11.080>
- Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C. P. E., van Dijk, C. M., Tollenaar, R. A. E. M., Van Den Berg, D., & Laird, P. W. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genetics*, 44(1), 40–46. <https://doi.org/10.1038/ng.969>
- Biswas, S., & Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology*, 837, 8–24. <https://doi.org/10.1016/j.ejphar.2018.08.021>
- Bluemn, T., Schmitz, J., Chen, Y., Zheng, Y., Zhang, Y., Zheng, S., Burns, R., DeJong, J., Christiansen, L., Izaguirre-Carbonell, J., Wang, D., & Zhu, N. (2021). Arid2 regulates hematopoietic stem cell differentiation in normal hematopoiesis. *Experimental Hematology*, 94, 37–46. <https://doi.org/10.1016/j.exphem.2020.12.004>
- Bluemn, T., Schmitz, J., Zheng, Y., Burns, R., Zheng, S., DeJong, J., Christiansen, L., Arnold, O., Izaguirre-Carbonell, J., Wang, D., Deshpande, A. J., & Zhu, N. (2022). Differential roles of BAF and PBAF subunits, Arid1b and Arid2, in MLL-AF9 leukemogenesis. *Leukemia*, 36(4), 946–955. <https://doi.org/10.1038/s41375-021-01505-w>
- Boag, J. M., Beesley, A. H., Firth, M. J., Freitas, J. R., Ford, J., Brigstock, D. R., de Klerk, N. H., & Kees, U. R. (2007). High expression of connective tissue growth factor in pre-B acute lymphoblastic leukaemia. *British Journal of Haematology*, 138(6), 740–748. <https://doi.org/10.1111/j.1365-2141.2007.06739.x>
- Bohlander, S. K. (2005). ETV6: A versatile player in leukemogenesis. *Seminars in Cancer Biology*, 15(3), 162–174. <https://doi.org/10.1016/j.semcancer.2005.01.008>
- Brady, S. W., Roberts, K. G., Gu, Z., Shi, L., Pounds, S., Pei, D., Cheng, C., Dai, Y., Devidas, M., Qu, C., Hill, A. N., Payne-Turner, D., Ma, X., Iacobucci,

- I., Baviskar, P., Wei, L., Arunachalam, S., Hagiwara, K., Liu, Y., ... Mullighan, C. G. (2022). The genomic landscape of pediatric acute lymphoblastic leukemia. *Nature Genetics*, 54(9), 1376–1389. <https://doi.org/10.1038/s41588-022-01159-z>
- Brogna, S., & Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*, 16(2), 107–113. <https://doi.org/10.1038/nsmb.1550>
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, 25(18), 1915–1927. <https://doi.org/10.1101/gad.17446611>
- Carcamo, S., Nguyen, C. B., Grossi, E., Filipescu, D., Alpsoy, A., Dhiman, A., Sun, D., Narang, S., Imig, J., Martin, T. C., Parsons, R., Aifantis, I., Tsirigos, A., Aguirre-Ghiso, J. A., Dykhuizen, E. C., Hasson, D., & Bernstein, E. (2022). Altered BAF occupancy and transcription factor dynamics in PBAF-deficient melanoma. *Cell Reports*, 39(1), 110637. <https://doi.org/10.1016/j.celrep.2022.110637>
- Cavalli, G., & Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature*, 571(7766), 489–499. <https://doi.org/10.1038/s41586-019-1411-0>
- Centore, R. C., Sandoval, G. J., Soares, L. M. M., Kadoch, C., & Chan, H. M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. *Trends in Genetics*, 36(12), 936–950. <https://doi.org/10.1016/j.tig.2020.07.011>
- Chang, J. H., Poppe, M. M., Hua, C., Marcus, K. J., & Esiashvili, N. (2021). Acute lymphoblastic leukemia. *Pediatric Blood & Cancer*, 68(S2). <https://doi.org/10.1002/pbc.28371>
- Chatterjee, A., Rodger, E. J., & Eccles, M. R. (2018). Epigenetic drivers of tumorigenesis and cancer metastasis. *Seminars in Cancer Biology*, 51, 149–159. <https://doi.org/10.1016/j.semcancer.2017.08.004>
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14(1), 128. <https://doi.org/10.1186/1471-2105-14-128>
- Chen, X.-Y., Zhang, J., & Zhu, J.-S. (2019). The role of m6A RNA methylation in human cancer. *Molecular Cancer*, 18(1), 103. <https://doi.org/10.1186/s12943-019-1033-z>

- Chennamadhavuni, A., Lyengar, V., Mukkamalla, S. K. R., & Shimanovsky, A. (2023). *Leukemia*.
- Chuang, T.-D., Quintanilla, D., Boos, D., & Khorram, O. (2022). Differential Expression of Super-Enhancer-Associated Long Non-coding RNAs in Uterine Leiomyomas. *Reproductive Sciences*, 29(10), 2960–2976. <https://doi.org/10.1007/s43032-022-00981-4>
- Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18(7), 407–422. <https://doi.org/10.1038/nrm.2017.26>
- Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., Dinger, M. E., & Mattick, J. S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Research*, 22(5), 885–898. <https://doi.org/10.1101/gr.131037.111>
- Coccaro, N., Anelli, L., Zagaria, A., Specchia, G., & Albano, F. (2019). Next-Generation Sequencing in Acute Lymphoblastic Leukemia. *International Journal of Molecular Sciences*, 20(12), 2929. <https://doi.org/10.3390/ijms20122929>
- Corces, M. R., Granja, J. M., Shams, S., Louie, B. H., Seoane, J. A., Zhou, W., Silva, T. C., Groeneveld, C., Wong, C. K., Cho, S. W., Satpathy, A. T., Mumbach, M. R., Hoadley, K. A., Robertson, A. G., Sheffield, N. C., Felau, I., Castro, M. A. A., Berman, B. P., Staudt, L. M., ... Zhu, J. (2018). The chromatin accessibility landscape of primary human cancers. *Science*, 362(6413). <https://doi.org/10.1126/science.aav1898>
- Cruz-Miranda, G., Hidalgo-Miranda, A., Bárcenas-López, D., Núñez-Enríquez, J., Ramírez-Bello, J., Mejía-Aranguré, J., & Jiménez-Morales, S. (2019). Long Non-Coding RNA and Acute Leukemia. *International Journal of Molecular Sciences*, 20(3), 735. <https://doi.org/10.3390/ijms20030735>
- Cuadros, M., Andrades, Á., Coira, I. F., Baliñas, C., Rodríguez, M. I., Álvarez-Pérez, J. C., Peinado, P., Arenas, A. M., García, D. J., Jiménez, P., Camós, M., Jiménez-Velasco, A., & Medina, P. P. (2019). Expression of the long non-coding RNA TCL6 is associated with clinical outcome in pediatric B-cell acute lymphoblastic leukemia. *Blood Cancer Journal*, 9(12), 93. <https://doi.org/10.1038/s41408-019-0258-9>
- Cuadros, M., García, D. J., Andrades, A., Arenas, A. M., Coira, I. F., Baliñas-Gavira, C., Peinado, P., Rodríguez, M. I., Álvarez-Pérez, J. C., Ruiz-Cabello, F., Camós, M., Jiménez-Velasco, A., & Medina, P. P. (2020). LncRNA-mRNA Co-Expression Analysis Identifies AL133346.1/CCN2 as

- Biomarkers in Pediatric B-Cell Acute Lymphoblastic Leukemia. *Cancers*, 12(12), 3803. <https://doi.org/10.3390/cancers12123803>
- Cunningham, R. M., Walton, M. A., & Carter, P. M. (2018). The Major Causes of Death in Children and Adolescents in the United States. *New England Journal of Medicine*, 379(25), 2468–2475. <https://doi.org/10.1056/NEJMSr1804754>
- Datta, N., Chakraborty, S., Basu, M., & Ghosh, M. K. (2020). Tumor Suppressors Having Oncogenic Functions: The Double Agents. *Cells*, 10(1), 46. <https://doi.org/10.3390/cells10010046>
- de Castro, R. O., Previato, L., Goitea, V., Felberg, A., Guiraldelli, M. F., Filiber-ti, A., & Pezza, R. J. (2017). The chromatin-remodeling subunit Baf200 promotes homology-directed DNA repair and regulates distinct chromatin-remodeling complexes. *Journal of Biological Chemistry*, 292(20), 8459–8471. <https://doi.org/10.1074/jbc.M117.778183>
- Dombret, H., Gabert, J., Boiron, J.-M., Rigal-Huguet, F., Blaise, D., Thomas, X., Delannoy, A., Buzyn, A., Bilhou-Nabera, C., Cayuela, J.-M., Fenaux, P., Bourhis, J.-H., Fegueux, N., Charrin, C., Boucheix, C., Lhéritier, V., Espérou, H., MacIntyre, E., Vernant, J.-P., ... Groupe d'Etude et de Traitement de la Leucémie Aiguë Lymphoblastique de l'Adulte (GET-LALA Group). (2002). Outcome of treatment in adults with Philadelphia chromosome-positive acute lymphoblastic leukemia--results of the prospective multicenter LALA-94 trial. *Blood*, 100(7), 2357–2366. <https://doi.org/10.1182/blood-2002-03-0704>
- Duan, Y., Tian, L., Gao, Q., Liang, L., Zhang, W., Yang, Y., Zheng, Y., Pan, E., Li, S., & Tang, N. (2016). Chromatin remodeling gene *ARID2* targets cyclin D1 and cyclin E1 to suppress hepatoma cell progression. *Oncotarget*, 7(29), 45863–45875. <https://doi.org/10.18632/oncotarget.10244>
- Duque-Afonso, J., Feng, J., Scherer, F., Lin, C.-H., Wong, S. H. K., Wang, Z., Iwasaki, M., & Cleary, M. L. (2015). Comparative genomics reveals multis-tep pathogenesis of E2A-PBX1 acute lymphoblastic leukemia. *Journal of Clinical Investigation*, 125(9), 3667–3680. <https://doi.org/10.1172/JCI81158>
- Dykes, I. M., & Emanuelli, C. (2017). Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics, Proteomics and Bioinformatics*, 15(3), 177–186. <https://doi.org/10.1016/J.GPB.2016.12.005>
- Dykhuizen, E. C., Hargreaves, D. C., Miller, E. L., Cui, K., Korshunov, A., Kool, M., Pfister, S., Cho, Y.-J., Zhao, K., & Crabtree, G. R. (2013). BAF com-

- plexes facilitate decatenation of DNA by topoisomerase IIa. *Nature*, 497(7451), 624–627. <https://doi.org/10.1038/nature12146>
- Ebralidze, A. K., Guibal, F. C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M. A., Petkova, V., Rosenbauer, F., Huang, G., Dayaram, T., Klupp, J., O'Brien, K. B., Will, B., Hoogenkamp, M., Borden, K. L. B., Bonifer, C., & Tenen, D. G. (2008). *PU.1* expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared *cis*-regulatory element. *Genes & Development*, 22(15), 2085–2092. <https://doi.org/10.1101/gad.1654808>
- Eccleston, A., DeWitt, N., Gunter, C., Marte, B., & Nath, D. (2007). Epigenetics. *Nature*, 447(7143), 395–395. <https://doi.org/10.1038/447395a>
- El Chaer, F., Keng, M., & Ballen, K. K. (2020). MLL-Rearranged Acute Lymphoblastic Leukemia. *Current Hematologic Malignancy Reports*, 15(2), 83–89. <https://doi.org/10.1007/s11899-020-00582-5>
- Fernando, T. R., Rodriguez-Malave, N. I., Waters, E. V., Yan, W., Casero, D., Basso, G., Pigazzi, M., & Rao, D. S. (2015). LncRNA Expression Discriminates Karyotype and Predicts Survival in B-Lymphoblastic Leukemia. *Molecular Cancer Research*, 13(5), 839–851. <https://doi.org/10.1158/1541-7786.MCR-15-0006-T>
- Fillmore, C. M., Xu, C., Desai, P. T., Berry, J. M., Rowbotham, S. P., Lin, Y.-J., Zhang, H., Marquez, V. E., Hammerman, P. S., Wong, K.-K., & Kim, C. F. (2015). EZH2 inhibition sensitizes BRG1 and EGFR mutant lung tumours to TopoII inhibitors. *Nature*, 520(7546), 239–242. <https://doi.org/10.1038/nature14122>
- Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T. I., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., & Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database : The Journal of Biological Databases and Curation*, 2017. <https://doi.org/10.1093/DATABASE/BAX028>
- Fox, A. H., Nakagawa, S., Hirose, T., & Bond, C. S. (2018). Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends in Biochemical Sciences*, 43(2), 124–135. <https://doi.org/10.1016/j.tibs.2017.12.001>
- Gao, J., Wang, F., Wu, P., Chen, Y., & Jia, Y. (2020). Aberrant LncRNA Expression in Leukemia. *Journal of Cancer*, 11(14), 4284–4296. <https://doi.org/10.7150/jca.42093>
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: evalua-

- ting next-generation sequencing alignment data. *Bioinformatics*, 28(20), 2678–2679. <https://doi.org/10.1093/bioinformatics/bts503>
- Gebuhr, T. C., Kovalev, G. I., Bultman, S., Godfrey, V., Su, L., & Magnuson, T. (2003). The Role of *Brg1*, a Catalytic Subunit of Mammalian Chromatin-remodeling Complexes, in T Cell Development. *The Journal of Experimental Medicine*, 198(12), 1937–1949. <https://doi.org/10.1084/jem.20030714>
- Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature Reviews Molecular Cell Biology*, 14(11), 699–712. <https://doi.org/10.1038/nrm3679>
- Giles, K. A., Gould, C. M., Achinger-Kawecka, J., Page, S. G., Kafer, G. R., Rogers, S., Luu, P.-L., Cesare, A. J., Clark, S. J., & Taberlay, P. C. (2021). BRG1 knockdown inhibits proliferation through multiple cellular pathways in prostate cancer. *Clinical Epigenetics*, 13(1), 37. <https://doi.org/10.1186/s13148-021-01023-7>
- Gingeras, T. R. (2007). Origin of phenotypes: genes and transcripts. *Genome Research*, 17(6), 682–690. <https://doi.org/10.1101/gr.6525007>
- Gopalakrishnan, S., Van Emburgh, B. O., & Robertson, K. D. (2008). DNA methylation in development and human disease. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 647(1–2), 30–38. <https://doi.org/10.1016/j.mrfmmm.2008.08.006>
- Górecki, M., Kozioł, I., Kopystecka, A., Budzyńska, J., Zawitkowska, J., & Lejman, M. (2023). Updates in KMT2A Gene Rearrangement in Pediatric Acute Lymphoblastic Leukemia. *Biomedicines*, 11(3), 821. <https://doi.org/10.3390/biomedicines11030821>
- Grabher, C., von Boehmer, H., & Look, A. T. (2006). Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. *Nature Reviews Cancer*, 6(5), 347–359. <https://doi.org/10.1038/nrc1880>
- Greaves, M. (2006). Infection, immune responses and the aetiology of childhood leukaemia. *Nature Reviews Cancer*, 6(3), 193–203. <https://doi.org/10.1038/nrc1816>
- Greaves, M., Cazzaniga, V., & Ford, • Anthony. (2021). Can we prevent childhood Leukaemia? Why prevention? *Leukemia*, 35, 1258–1264. <https://doi.org/10.1038/s41375-021-01211-7>
- Grossmann, V., Kern, W., Harbich, S., Alpermann, T., Jeromin, S., Schnittger, S., Haferlach, C., Haferlach, T., & Kohlmann, A. (2011). Prognostic relevance of RUNX1 mutations in T-cell acute lymphoblastic leukemia. *Hae-*

- matologica*, 96(12), 1874–1877.  
<https://doi.org/10.3324/haematol.2011.043919>
- 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>
- Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G., Yang, L., & Chen, L.-L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, 181(3), 621–636.e22. <https://doi.org/10.1016/j.cell.2020.03.006>
- Guo, Y., Zhao, S., Sheng, Q., Guo, M., Lehmann, B., Pietenpol, J., Samuels, D. C., & Shyr, Y. (2015). RNAseq by Total RNA Library Identifies Additional RNAs Compared to Poly(A) RNA Library. *BioMed Research International*, 2015, 1–9. <https://doi.org/10.1155/2015/862130>
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J. L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R. B., van de Vijver, M. J., Sukumar, S., & Chang, H. Y. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464(7291), 1071–1076. <https://doi.org/10.1038/nature08975>
- Gutschner, T., Hämmerle, M., & Diederichs, S. (2013). MALAT1 — a paradigm for long noncoding RNA function in cancer. *Journal of Molecular Medicine*, 91(7), 791–801. <https://doi.org/10.1007/s00109-013-1028-y>
- Haferlach, T., Kohlmann, A., Schnittger, S., Dugas, M., Hiddemann, W., Kern, W., & Schoch, C. (2005). Global approach to the diagnosis of leukemia using gene expression profiling. *Blood*, 106(4), 1189–1198. <https://doi.org/10.1182/blood-2004-12-4938>
- Han, L., Madan, V., Mayakonda, A., Dakle, P., Woon, T. W., Shyamsunder, P., Nordin, H. B. M., Cao, Z., Sundaresan, J., Lei, I., Wang, Z., & Koeffler, H. P. (2019). Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. *Leukemia*, 33(9), 2291–2305. <https://doi.org/10.1038/s41375-019-0438-4>
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>

- Hanly, D. J., Esteller, M., & Berdasco, M. (2018). Interplay between long non-coding RNAs and epigenetic machinery: emerging targets in cancer? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1748), 20170074. <https://doi.org/10.1098/rstb.2017.0074>
- Harrison, C. J., Moorman, A. V., Schwab, C., Iacobucci, I., & Mullighan, C. (2017). Cytogenetics and Molecular Genetics. In *Childhood Acute Lymphoblastic Leukemia* (pp. 61–98). Springer International Publishing. [https://doi.org/10.1007/978-3-319-39708-5\\_4](https://doi.org/10.1007/978-3-319-39708-5_4)
- Hart, S. N., Cui, Y., Klaassen, C. D., & Zhong, X. (2009). Three Patterns of Cytochrome P450 Gene Expression during Liver Maturation in Mice. *Drug Metabolism and Disposition*, 37(1), 116–121. <https://doi.org/10.1124/dmd.108.023812>
- Hashiguchi, S., Tanaka, T., Mano, R., Kondo, S., & Kodama, S. (2022). CCN2-induced lymphangiogenesis is mediated by the integrin  $\alpha\beta 5$ –ERK pathway and regulated by DUSP6. *Scientific Reports*, 12(1), 926. <https://doi.org/10.1038/s41598-022-04988-4>
- He, L., Tian, X., Zhang, H., Hu, T., Huang, X., Zhang, L., Wang, Z., & Zhou, B. (2014). BAF200 Is Required for Heart Morphogenesis and Coronary Artery Development. *PLoS ONE*, 9(10), e109493. <https://doi.org/10.1371/journal.pone.0109493>
- Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., Kim, Y., Kryukov, G. V., Ghandi, M., Aguirre, A. J., Jagani, Z., Wang, Z., Garraway, L. A., Hahn, W. C., & Roberts, C. W. M. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature Medicine*, 20(3), 251–254. <https://doi.org/10.1038/nm.3480>
- Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Uliatsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports*, 11(7), 1110–1122. <https://doi.org/10.1016/j.celrep.2015.04.023>
- Hiramatsu, Y., Fukuda, A., Ogawa, S., Goto, N., Ikuta, K., Tsuda, M., Matsumoto, Y., Kimura, Y., Yoshioka, T., Takada, Y., Maruno, T., Hanyu, Y., Tsuryama, T., Wang, Z., Akiyama, H., Takaishi, S., Miyoshi, H., Taketo, M. M., Chiba, T., & Seno, H. (2019). Arid1a is essential for intestinal stem cells through Sox9 regulation. *Proceedings of the National Academy of Sciences*, 116(5), 1704–1713. <https://doi.org/10.1073/pnas.1804858116>
- Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J.-P., Nickerson, E., Auclair, D., Li, L., Place, C., DiCara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G.,

- Stransky, N., Onofrio, R. C., ... Chin, L. (2012). A Landscape of Driver Mutations in Melanoma. *Cell*, 150(2), 251–263. <https://doi.org/10.1016/j.cell.2012.06.024>
- Hohmann, A. F., & Vakoc, C. R. (2014). A rationale to target the SWI/SNF complex for cancer therapy. *Trends in Genetics*, 30(8), 356–363. <https://doi.org/10.1016/j.tig.2014.05.001>
- Huang, J., Chan, S. C., Ngai, C. H., Lok, V., Zhang, L., Lucero-Prisno, D. E., Xu, W., Zheng, Z., Elcarte, E., Withers, M., & Wong, M. C. S. (2023). Global incidence, mortality and temporal trends of cancer in children: A joinpoint regression analysis. *Cancer Medicine*, 12(2), 1903–1911. <https://doi.org/10.1002/cam4.5009>
- Hui, X., Zhu, W., Wang, Y., Lam, K. S. L., Zhang, J., Wu, D., Kraegen, E. W., Li, Y., & Xu, A. (2009). Major Urinary Protein-1 Increases Energy Expenditure and Improves Glucose Intolerance through Enhancing Mitochondrial Function in Skeletal Muscle of Diabetic Mice. *Journal of Biological Chemistry*, 284(21), 14050–14057. <https://doi.org/10.1074/jbc.M109.001107>
- Hunger, S. P., & Mullighan, C. G. (2015a). Acute Lymphoblastic Leukemia in Children. *New England Journal of Medicine*, 373(16), 1541–1552. <https://doi.org/10.1056/NEJMra1400972>
- Hunger, S. P., & Mullighan, C. G. (2015b). Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood*, 125(26), 3977–3987. <https://doi.org/10.1182/blood-2015-02-580043>
- Inaba, H., Azzato, E. M., & Mullighan, C. G. (2017). Integration of Next-Generation Sequencing to Treat Acute Lymphoblastic Leukemia with Targetable Lesions: The St. Jude Children’s Research Hospital Approach. *Frontiers in Pediatrics*, 5. <https://doi.org/10.3389/fped.2017.00258>
- Inaba, H., Greaves, M., & Mullighan, C. G. (2013). Acute lymphoblastic leukaemia. *The Lancet*, 381(9881), 1943–1955. [https://doi.org/10.1016/S0140-6736\(12\)62187-4](https://doi.org/10.1016/S0140-6736(12)62187-4)
- Inaba, H., & Mullighan, C. G. (2020). Pediatric acute lymphoblastic leukemia. *Haematologica*, 105(11), 2524–2539. <https://doi.org/10.3324/haematol.2020.247031>
- Isidro-Hernández, M., Alemán-Arteaga, S., Casado-García, A., Ruiz-Corzo, B., Riesco, S., Prieto-Matos, P., Martínez-Cano, J., Sánchez, L., Cobaleda, C., Sánchez-García, I., & Vicente-Dueñas, C. (2022). Childhood B-Cell Pre-leukemia Mouse Modeling. *International Journal of Molecular Sciences*, 23(14), 7562. <https://doi.org/10.3390/ijms23147562>

- Ivkovic, S., Yoon, B. S., Popoff, S. N., Safadi, F. F., Libuda, D. E., Stephenson, R. C., Daluiski, A., & Lyons, K. M. (2003). Connective tissue growth factor coordinates chondrogenesis and angiogenesis during skeletal development. *Development*, *130*(12), 2779–2791. <https://doi.org/10.1242/dev.00505>
- Jackson, T. R., Ling, R. E., & Roy, A. (2021). The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia. *Frontiers in Immunology*, *12*. <https://doi.org/10.3389/fimmu.2021.637975>
- Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*(S3), 245–254. <https://doi.org/10.1038/ng1089>
- Jamieson, C. H. M., Amylon, M. D., Wong, R. M., & Blume, K. G. (2003). Allogeneic hematopoietic cell transplantation for patients with high-risk acute lymphoblastic leukemia in first or second complete remission using fractionated total-body irradiation and high-dose etoposide: a 15-year experience. *Experimental Hematology*, *31*(10), 981–986. [https://doi.org/10.1016/s0301-472x\(03\)00231-5](https://doi.org/10.1016/s0301-472x(03)00231-5)
- Jarroux, J., Morillon, A., & Pinskaya, M. (2017). *History, Discovery, and Classification of lncRNAs* (pp. 1–46). [https://doi.org/10.1007/978-981-10-5203-3\\_1](https://doi.org/10.1007/978-981-10-5203-3_1)
- Jassim, A., Rahrman, E. P., Simons, B. D., & Gilbertson, R. J. (2023). Cancers make their own luck: theories of cancer origins. *Nature Reviews Cancer*. <https://doi.org/10.1038/s41568-023-00602-5>
- Jędraszek, K., Malczewska, M., Parysek-Wójcik, K., & Lejman, M. (2022). Resistance Mechanisms in Pediatric B-Cell Acute Lymphoblastic Leukemia. *International Journal of Molecular Sciences*, *23*(6), 23. <https://doi.org/10.3390/IJMS23063067>
- Jiang, Z., Peng, T., Qian, H., Lu, C., Qiu, F., & Zhang, S. (2019). DNA damage-induced activation of ATM promotes  $\beta$ -TRCP-mediated ARID1A ubiquitination and destruction in gastric cancer cells. *Cancer Cell International*, *19*(1), 162. <https://doi.org/10.1186/s12935-019-0878-y>
- Jones, C. A., Tansey, W. P., & Weissmiller, A. M. (2022). Emerging Themes in Mechanisms of Tumorigenesis by SWI/SNF Subunit Mutation. *Epigenetics Insights*, *15*, 25168657221115656. <https://doi.org/10.1177/25168657221115656>
- Kaczmarek, A., Derebas, J., Pinkosz, M., Niedźwiecki, M., & Lejman, M. (2023). The Landscape of Secondary Genetic Rearrangements in Pediatric

- Patients with B-Cell Acute Lymphoblastic Leukemia with t(12;21). *Cells*, 12(3), 357. <https://doi.org/10.3390/cells12030357>
- Kamieniarz-Gdula, K., & Proudfoot, N. J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. *Trends in Genetics*, 35(8), 553–564. <https://doi.org/10.1016/j.tig.2019.05.005>
- Kamori, T., Oki, E., Shimada, Y., Hu, Q., Hisamatsu, Y., Ando, K., Shimokawa, M., Wakai, T., Oda, Y., & Mori, M. (2022). The effects of *ARID1A* mutations on colorectal cancer and associations with *PD-L1* expression by stromal cells. *Cancer Reports*, 5(1). <https://doi.org/10.1002/cnr2.1420>
- Kang, H., Chen, I.-M., Wilson, C. S., Bedrick, E. J., Harvey, R. C., Atlas, S. R., Devidas, M., Mullighan, C. G., Wang, X., Murphy, M., Ar, K., Wharton, W., Borowitz, M. J., Bowman, W. P., Bhojwani, D., Carroll, W. L., Camitta, B. M., Reaman, G. H., Smith, M. A., ... Willman, C. L. (2010). Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood*, 115(7), 1394–1405. <https://doi.org/10.1182/blood-2009-05-218560>
- Karatas, H., Townsend, E. C., Cao, F., Chen, Y., Bernard, D., Liu, L., Lei, M., Dou, Y., & Wang, S. (2013). High-Affinity, Small-Molecule Peptidomimetic Inhibitors of MLL1/WDR5 Protein–Protein Interaction. *Journal of the American Chemical Society*, 135(2), 669–682. <https://doi.org/10.1021/ja306028q>
- Katlinskaya, Y. V., Katlinski, K. V., Lasri, A., Li, N., Beiting, D. P., Durham, A. C., Yang, T., Pikarsky, E., Lengner, C. J., Johnson, F. B., Ben-Neriah, Y., & Fuchs, S. Y. (2016). Type I Interferons Control Proliferation and Function of the Intestinal Epithelium. *Molecular and Cellular Biology*, 36(7), 1124–1135. <https://doi.org/10.1128/MCB.00988-15>
- Ke, S., Alemu, E. A., Mertens, C., Gantman, E. C., Fak, J. J., Mele, A., Haripal, B., Zucker-Scharff, I., Moore, M. J., Park, C. Y., Vågbø, C. B., Kusnierczyk, A., Klungland, A., Darnell, J. E., & Darnell, R. B. (2015). A majority of m<sup>6</sup>A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes & Development*, 29(19), 2037–2053. <https://doi.org/10.1101/gad.269415.115>
- Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., & Rinn, J. L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene

- expression. *Proceedings of the National Academy of Sciences*, 106(28), 11667–11672. <https://doi.org/10.1073/pnas.0904715106>
- Kharel, A., Shen, J., Brown, R., Chen, Y., Nguyen, C., Alson, D., Bluemn, T., Fan, J., Gai, K., Zhang, B., Kudek, M., Zhu, N., & Cui, W. (2023). Loss of PBAF promotes expansion and effector differentiation of CD8+ T cells during chronic viral infection and cancer. *Cell Reports*, 42(6), 112649. <https://doi.org/10.1016/j.celrep.2023.112649>
- Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, 172(3), 393–407. <https://doi.org/10.1016/J.CELL.2018.01.011>
- Kubota, S., & Takigawa, M. (2015). Cellular and molecular actions of CCN2/CTGF and its role under physiological and pathological conditions. *Clinical Science*, 128(3), 181–196. <https://doi.org/10.1042/CS20140264>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97. <https://doi.org/10.1093/nar/gkw377>
- Kurdistani, S. K. (2011). Histone Modifications in Cancer Biology and Prognosis. In *Epigenetics and Disease* (pp. 91–106). Springer Basel. [https://doi.org/10.1007/978-3-7643-8989-5\\_5](https://doi.org/10.1007/978-3-7643-8989-5_5)
- Laird-Offringa, I. A., & Sanchez-Cespedes, M. (2018). Epigenetic Events in Lung Cancer. In *IASLC Thoracic Oncology* (pp. 104-116.e5). Elsevier. <https://doi.org/10.1016/B978-0-323-52357-8.00012-3>
- Law, G. R., Parslow, R. C., & Roman, E. (2003). Childhood Cancer and Population Mixing. *American Journal of Epidemiology*, 158(4), 328–336. <https://doi.org/10.1093/aje/kwg165>
- Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., & Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, 505(7484), 495–501. <https://doi.org/10.1038/nature12912>
- Lebedev, T., Kousar, R., Patrick, B., Usama, M., Lee, M.-K., Tan, M., & Li, X.-G. (2023). Targeting ARID1A-Deficient Cancers: An Immune-Metabolic Perspective. *Cells*, 12(6), 952. <https://doi.org/10.3390/cells12060952>
- Lee, H., Dai, F., Zhuang, L., Xiao, Z.-D., Kim, J., Zhang, Y., Ma, L., You, M. J., Wang, Z., & Gan, B. (2016). BAF180 regulates cellular senescence and

- hematopoietic stem cell homeostasis through p21. *Oncotarget*, 7(15), 19134–19146. <https://doi.org/10.18632/oncotarget.8102>
- Leguit, R. J., Raymakers, R. A. P., Hebeda, K. M., & Goldschmeding, R. (2021). CCN2 (Cellular Communication Network factor 2) in the bone marrow microenvironment, normal and malignant hematopoiesis. *Journal of Cell Communication and Signaling*, 15(1), 25–56. <https://doi.org/10.1007/s12079-020-00602-2>
- Lenz, G., Wright, G., Dave, S. S., Xiao, W., Powell, J., Zhao, H., Xu, W., Tan, B., Goldschmidt, N., Iqbal, J., Vose, J., Bast, M., Fu, K., Weisenburger, D. D., Greiner, T. C., Armitage, J. O., Kyle, A., May, L., Gascoyne, R. D., ... Staudt, L. M. (2008). Stromal Gene Signatures in Large-B-Cell Lymphomas. *New England Journal of Medicine*, 359(22), 2313–2323. <https://doi.org/10.1056/NEJMoa0802885>
- Li, S., Li, Y., Chen, B., Zhao, J., Yu, S., Tang, Y., Zheng, Q., Li, Y., Wang, P., He, X., & Huang, S. (2018). exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Research*, 46(D1), D106–D112. <https://doi.org/10.1093/nar/gkx891>
- Li, X. S., Trojer, P., Matsumura, T., Treisman, J. E., & Tanese, N. (2010). Mammalian SWI/SNF-A Subunit BAF250/ARID1 Is an E3 Ubiquitin Ligase That Targets Histone H2B. *Molecular and Cellular Biology*, 30(7), 1673–1688. <https://doi.org/10.1128/MCB.00540-09>
- Li, Z., Zhao, J., & Tang, Y. (2023). Advances in the role of <scp>SWI</scp> / <scp>SNF</scp> complexes in tumours. *Journal of Cellular and Molecular Medicine*, 27(8), 1023–1031. <https://doi.org/10.1111/jcmm.17709>
- Liao, Q., Liu, C., Yuan, X., Kang, S., Miao, R., Xiao, H., Zhao, G., Luo, H., Bu, D., Zhao, H., Skogerbø, G., Wu, Z., & Zhao, Y. (2011). Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. *Nucleic Acids Research*, 39(9), 3864–3878. <https://doi.org/10.1093/nar/gkq1348>
- Lindeboom, R. G. H., Supek, F., & Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nature Genetics*, 48(10), 1112–1118. <https://doi.org/10.1038/ng.3664>
- Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315–322. <https://doi.org/10.1038/nature08514>

- Liu, L., Wan, X., Zhou, P., Zhou, X., Zhang, W., Hui, X., Yuan, X., Ding, X., Zhu, R., Meng, G., Xiao, H., Ma, F., Huang, H., Song, X., Zhou, B., Xiong, S., & Zhang, Y. (2018). The chromatin remodeling subunit Baf200 promotes normal hematopoiesis and inhibits leukemogenesis. *Journal of Hematology & Oncology*, *11*(1), 27. <https://doi.org/10.1186/s13045-018-0567-7>
- Liu, M., Sun, T., Li, N., Peng, J., Fu, D., Li, W., Li, L., & Gao, W.-Q. (2019). BRG1 attenuates colonic inflammation and tumorigenesis through autophagy-dependent oxidative stress sequestration. *Nature Communications*, *10*(1), 4614. <https://doi.org/10.1038/s41467-019-12573-z>
- Liu, Y., Huang, M., Wang, X., Liu, Z., Li, S., & Chen, Y.-G. (2023). Segregation of the stemness program from the proliferation program in intestinal stem cells. *Stem Cell Reports*, *18*(5), 1196–1210. <https://doi.org/10.1016/j.stemcr.2023.03.007>
- Liu-Dumlao, T., Kantarjian, H., Thomas, D. A., O'Brien, S., & Ravandi, F. (2012). Philadelphia-Positive Acute Lymphoblastic Leukemia: Current Treatment Options. *Current Oncology Reports*, *14*(5), 387–394. <https://doi.org/10.1007/s11912-012-0247-7>
- Lobry, C., Ntziachristos, P., Ndiaye-Lobry, D., Oh, P., Cimmino, L., Zhu, N., Araldi, E., Hu, W., Freund, J., Abdel-Wahab, O., Ibrahim, S., Skokos, D., Armstrong, S. A., Levine, R. L., Park, C. Y., & Aifantis, I. (2013). Notch pathway activation targets AML-initiating cell homeostasis and differentiation. *Journal of Experimental Medicine*, *210*(2), 301–319. <https://doi.org/10.1084/jem.20121484>
- Loesch, R., Chenane, L., & Colnot, S. (2020). ARID2 Chromatin Remodeler in Hepatocellular Carcinoma. *Cells*, *9*(10), 2152. <https://doi.org/10.3390/cells9102152>
- Long, Y., Wang, X., Youmans, D. T., & Cech, T. R. (2017). How do lncRNAs regulate transcription? *Science Advances*, *3*(9). <https://doi.org/10.1126/sciadv.aao2110>
- Lou, X., Zhang, J., Liu, S., Lou, X., & Liao, D. J. (2014). The other side of the coin: The tumor-suppressive aspect of oncogenes and the oncogenic aspect of tumor-suppressive genes, such as those along the CCND–CDK4/6–RB axis. *Cell Cycle*, *13*(11), 1677–1693. <https://doi.org/10.4161/cc.29082>
- Lu, Y., Chan, Y.-T., Tan, H.-Y., Li, S., Wang, N., & Feng, Y. (2020). Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. *Molecular Cancer*, *19*(1), 79. <https://doi.org/10.1186/s12943-020-01197-3>

- Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biology*, 10(6), 925–933. <https://doi.org/10.4161/rna.24604>
- Ma, X., Edmonson, M., Yergeau, D., Muzny, D. M., Hampton, O. A., Rusch, M., Song, G., Easton, J., Harvey, R. C., Wheeler, D. A., Ma, J., Doddapaneni, H., Vadodaria, B., Wu, G., Nagahawatte, P., Carroll, W. L., Chen, I.-M., Gastier-Foster, J. M., Relling, M. V., ... Zhang, J. (2015). Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nature Communications*, 6(1), 6604. <https://doi.org/10.1038/ncomms7604>
- Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., Zhou, X., Li, Y., Rusch, M. C., Easton, J., Huether, R., Gonzalez-Pena, V., Wilkinson, M. R., Hermida, L. C., Davis, S., Sioson, E., Pounds, S., Cao, X., Ries, R. E., ... Zhang, J. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*, 555(7696), 371–376. <https://doi.org/10.1038/nature25795>
- Magaña-Acosta, M., & Valadez-Graham, V. (2020). Chromatin Remodelers in the 3D Nuclear Compartment. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.600615>
- Malard, F., & Mohty, M. (2020). Acute lymphoblastic leukaemia. *Lancet (London, England)*, 395(10230), 1146–1162. [https://doi.org/10.1016/S0140-6736\(19\)33018-1](https://doi.org/10.1016/S0140-6736(19)33018-1)
- Malczewska, M., Kośmider, K., Bednarz, K., Ostapińska, K., Lejman, M., & Zawitkowska, J. (2022). Recent Advances in Treatment Options for Childhood Acute Lymphoblastic Leukemia. *Cancers*, 14(8), 2021. <https://doi.org/10.3390/cancers14082021>
- Manceau, G., Letouzé, E., Guichard, C., Didelot, A., Cazes, A., Corté, H., Fabre, E., Pallier, K., Imbeaud, S., Le Pimpec-Barthes, F., Zucman-Rossi, J., Laurent-Puig, P., & Blons, H. (2013). Recurrent inactivating mutations of *ARID2* in non-small cell lung carcinoma. *International Journal of Cancer*, 132(9), 2217–2221. <https://doi.org/10.1002/ijc.27900>
- Marcotte, E. L., Spector, L. G., Mendes-de-Almeida, D. P., & Nelson, H. H. (2021). The Prenatal Origin of Childhood Leukemia: Potential Applications for Epidemiology and Newborn Screening. *Frontiers in Pediatrics*, 9. <https://doi.org/10.3389/fped.2021.639479>
- Martelli, A. M., Paganelli, F., Fazio, A., Bazzichetto, C., Conciatori, F., & McCubrey, J. A. (2019). The Key Roles of PTEN in T-Cell Acute

- Lymphoblastic Leukemia Development, Progression, and Therapeutic Response. *Cancers*, 11(5), 629. <https://doi.org/10.3390/cancers11050629>
- Maru, G. B. (2016). Understanding the molecular mechanisms of cancer prevention by dietary phytochemicals: From experimental models to clinical trials. *World Journal of Biological Chemistry*, 7(1), 88. <https://doi.org/10.4331/wjbc.v7.i1.88>
- Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullo, H. J., McKenzie, Z. M., Kubiak, R. L., St. Pierre, R., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, 175(5), 1272-1288.e20. <https://doi.org/10.1016/j.cell.2018.09.032>
- Maslah-Planchon, J., Bièche, I., Guinebretière, J.-M., Bourdeaut, F., & Delattre, O. (2015). SWI/SNF Chromatin Remodeling and Human Malignancies. *Annual Review of Pathology: Mechanisms of Disease*, 10(1), 145–171. <https://doi.org/10.1146/annurev-pathol-012414-040445>
- Mathur, R., Alver, B. H., San Roman, A. K., Wilson, B. G., Wang, X., Agoston, A. T., Park, P. J., Shivdasani, R. A., & Roberts, C. W. M. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nature Genetics*, 49(2), 296–302. <https://doi.org/10.1038/ng.3744>
- Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L.-L., Chen, R., Dean, C., Dinger, M. E., Fitzgerald, K. A., Gingeras, T. R., Guttman, M., Hirose, T., Huarte, M., Johnson, R., Kanduri, C., Kapranov, P., Lawrence, J. B., Lee, J. T., ... Wu, M. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature Reviews Molecular Cell Biology*. <https://doi.org/10.1038/s41580-022-00566-8>
- McFarland, C. D., Yaglom, J. A., Wojtkowiak, J. W., Scott, J. G., Morse, D. L., Sherman, M. Y., & Mirny, L. A. (2017). The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Research*, 77(18), 4763–4772. <https://doi.org/10.1158/0008-5472.CAN-15-3283-T>
- Melé, M., Mattioli, K., Mallard, W., Shechner, D. M., Gerhardinger, C., & Rinn, J. L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Research*, 27(1), 27–37. <https://doi.org/10.1101/gr.214205.116>
- Menon, D. U., Kirsanov, O., Geyer, C. B., & Magnuson, T. (2021). Mammalian SWI/SNF chromatin remodeler is essential for reductional meiosis in males. *Nature Communications*, 12(1), 6581. <https://doi.org/10.1038/s41467-021-26828-1>

- Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E., & Jaffrey, S. R. (2012). Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell*, *149*(7), 1635–1646. <https://doi.org/10.1016/j.cell.2012.05.003>
- Mittal, P., & Roberts, C. W. M. (2020). The SWI/SNF complex in cancer — biology, biomarkers and therapy. *Nature Reviews Clinical Oncology*, *17*(7), 435–448. <https://doi.org/10.1038/s41571-020-0357-3>
- Montefiori, L. E., Bendig, S., Gu, Z., Chen, X., Pölönen, P., Ma, X., Murison, A., Zeng, A., Garcia-Prat, L., Dickerson, K., Iacobucci, I., Abdelhamed, S., Hiltenbrand, R., Mead, P. E., Mehr, C. M., Xu, B., Cheng, Z., Chang, T.-C., Westover, T., ... Mullighan, C. G. (2021). Enhancer Hijacking Drives Oncogenic *BCL11B* Expression in Lineage-Ambiguous Stem Cell Leukemia. *Cancer Discovery*, *11*(11), 2846–2867. <https://doi.org/10.1158/2159-8290.CD-21-0145>
- Mullighan, C. G. (2014). The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology*, *2014*(1), 174–180. <https://doi.org/10.1182/asheducation-2014.1.174>
- Mullighan, C. G., Jeha, S., Pei, D., Payne-Turner, D., Coustan-Smith, E., Roberts, K. G., Waanders, E., Choi, J. K., Ma, X., Raimondi, S. C., Fan, Y., Yang, W., Song, G., Yang, J. J., Inaba, H., Downing, J. R., Leung, W. H., Bowman, W. P., Relling, M. V., ... Pui, C.-H. (2015). Outcome of children with hypodiploid ALL treated with risk-directed therapy based on MRD levels. *Blood*, *126*(26), 2896–2899. <https://doi.org/10.1182/blood-2015-09-671131>
- Mullighan, C. G., Phillips, L. A., Su, X., Ma, J., Miller, C. B., Shurtleff, S. A., & Downing, J. R. (2008). Genomic Analysis of the Clonal Origins of Relapsed Acute Lymphoblastic Leukemia. *Science*, *322*(5906), 1377–1380. <https://doi.org/10.1126/science.1164266>
- Nagamachi, A., Yamasaki, N., Miyazaki, K., Oda, H., Miyazaki, M., Honda, Z., Kominami, R., Inaba, T., & Honda, H. (2009). Haploinsufficiency and acquired loss of *Bcl11b* and *H2AX* induces blast crisis of chronic myelogenous leukemia in a transgenic mouse model. *Cancer Science*, *100*(7), 1219–1226. <https://doi.org/10.1111/j.1349-7006.2009.01172.x>
- Naidu, S. R., Capitano, M., Ropa, J., Cooper, S., Huang, X., & Broxmeyer, H. E. (2022). Chromatin remodeling subunit BRM and valine regulate hematopoietic stem/progenitor cell function and self-renewal via intrinsic and extrinsic effects. *Leukemia*, *36*(3), 821–833. <https://doi.org/10.1038/s41375-021-01426-8>

- Namayandeh, S. M., Khazaei, Z., Lari Najafi, M., Goodarzi, E., & Moslem, A. (2020). GLOBAL Leukemia in Children 0-14 Statistics 2018, Incidence and Mortality and Human Development Index (HDI): GLOBOCAN Sources and Methods. *Asian Pacific Journal of Cancer Prevention*, 21(5), 1487–1494. <https://doi.org/10.31557/APJCP.2020.21.5.1487>
- Nicoś, M., & Krawczyk, P. (2022). Genetic Clonality as the Hallmark Driving Evolution of Non-Small Cell Lung Cancer. *Cancers*, 14(7), 1813. <https://doi.org/10.3390/cancers14071813>
- Nie, Z., Yan, Z., Chen, E. H., Sechi, S., Ling, C., Zhou, S., Xue, Y., Yang, D., Murray, D., Kanakubo, E., Cleary, M. L., & Wang, W. (2003). Novel SWI/SNF Chromatin-Remodeling Complexes Contain a Mixed-Lineage Leukemia Chromosomal Translocation Partner. *Molecular and Cellular Biology*, 23(8), 2942–2952. <https://doi.org/10.1128/MCB.23.8.2942-2952.2003>
- Nojima, T., & Proudfoot, N. J. (2022). Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. *Nature Reviews Molecular Cell Biology*, 23(6), 389–406. <https://doi.org/10.1038/s41580-021-00447-6>
- Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S. M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., & Proudfoot, N. J. (2018). De-regulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Molecular Cell*, 72(6), 970-984.e7. <https://doi.org/10.1016/j.molcel.2018.10.011>
- Nussinov, R., & Tsai, C.-J. (2015). ‘Latent drivers’ expand the cancer mutational landscape. *Current Opinion in Structural Biology*, 32, 25–32. <https://doi.org/10.1016/j.sbi.2015.01.004>
- Oba, A., Shimada, S., Akiyama, Y., Nishikawaji, T., Mogushi, K., Ito, H., Matsumura, S., Aihara, A., Mitsunori, Y., Ban, D., Ochiai, T., Kudo, A., Asahara, H., Kaida, A., Miura, M., Tanabe, M., & Tanaka, S. (2017). ARID2 modulates DNA damage response in human hepatocellular carcinoma cells. *Journal of Hepatology*, 66(5), 942–951. <https://doi.org/10.1016/j.jhep.2016.12.026>
- Odnokoz, O., Wavelet-Vermuse, C., Hophan, S. L., Bulun, S., & Wan, Y. (2021). ARID1 proteins: from transcriptional and post-translational regulation to carcinogenesis and potential therapeutics. *Epigenomics*, 13(10), 809–823. <https://doi.org/10.2217/epi-2020-0414>
- Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R., & Shiekhattar,

- R. (2010). Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell*, 143(1), 46–58. <https://doi.org/10.1016/j.cell.2010.09.001>
- Otto, J. E., & Kadoch, C. (2017). A Two-Faced mSWI/SNF Subunit: Dual Roles for ARID1A in Tumor Suppression and Oncogenicity in the Liver. *Cancer Cell*, 32(5), 542–543. <https://doi.org/10.1016/j.ccell.2017.10.014>
- Pagliaroli, L., & Trizzino, M. (2021). The Evolutionary Conserved SWI/SNF Subunits ARID1A and ARID1B Are Key Modulators of Pluripotency and Cell-Fate Determination. *Frontiers in Cell and Developmental Biology*, 9. <https://doi.org/10.3389/fcell.2021.643361>
- Pang, D., McNally, R., & Birch, J. M. (2003). Parental smoking and childhood cancer : results from the United Kingdom Childhood Cancer Study. *British Journal of Cancer*, 88, 373–381. <https://doi.org/10.1038/sj.bjc.6600774>
- Pang, L., Huynh, J., Alorro, M. G., Li, X., Ernst, M., & Chand, A. L. (2021). STAT3 Signalling via the IL-6ST/gp130 Cytokine Receptor Promotes Epithelial Integrity and Intestinal Barrier Function during DSS-Induced Colitis. *Biomedicines*, 9(2), 187. <https://doi.org/10.3390/biomedicines9020187>
- Paralkar, V. R., Tabora, C. C., Huang, P., Yao, Y., Kossenkov, A. V., Prasad, R., Luan, J., Davies, J. O. J., Hughes, J. R., Hardison, R. C., Blobel, G. A., & Weiss, M. J. (2016). Unlinking an lncRNA from Its Associated cis Element. *Molecular Cell*, 62(1), 104–110. <https://doi.org/10.1016/J.MOLCEL.2016.02.029>
- Pathania, A. S., Prathipati, P., Pandey, M. K., Byrareddy, S. N., Coulter, D. W., Gupta, S. C., & Challagundla, K. B. (2022). The emerging role of non-coding RNAs in the epigenetic regulation of pediatric cancers. *Seminars in Cancer Biology*, 83, 227–241. <https://doi.org/10.1016/j.semcancer.2021.04.015>
- Paulsson, K., Lilljebjörn, H., Biloglav, A., Olsson, L., Rissler, M., Castor, A., Barbany, G., Fogelstrand, L., Nordgren, A., Sjögren, H., Fioretos, T., & Johansson, B. (2015). The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nature Publishing Group*, 47(6). <https://doi.org/10.1038/ng.3301>
- PDQ Pediatric Treatment Editorial Board. (2002). *Childhood Acute Lymphoblastic Leukemia Treatment (PDQ®): Health Professional Version*.
- Perbal, B. (2013). CCN proteins: A centralized communication network. *Journal of Cell Communication and Signaling*, 7(3), 169–177. <https://doi.org/10.1007/s12079-013-0193-7>

- Petri, A., Dybkær, K., Bøgsted, M., Thruø, C. A., Hagedorn, P. H., Schmitz, A., Bødker, J. S., Johnsen, H. E., & Kauppinen, S. (2015). Long Noncoding RNA Expression during Human B-Cell Development. *PLOS ONE*, *10*(9), e0138236. <https://doi.org/10.1371/journal.pone.0138236>
- Pikor, L., Thu, K., Vucic, E., & Lam, W. (2013). The detection and implication of genome instability in cancer. *Cancer and Metastasis Reviews*, *32*(3–4), 341–352. <https://doi.org/10.1007/s10555-013-9429-5>
- Priam, P., Krasteva, V., Rousseau, P., D'Angelo, G., Gaboury, L., Sauvageau, G., & Lessard, J. A. (2017). SMARCD2 subunit of SWI/SNF chromatin-remodeling complexes mediates granulopoiesis through a CEBP $\epsilon$  dependent mechanism. *Nature Genetics*, *49*(5), 753–764. <https://doi.org/10.1038/ng.3812>
- 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>
- Pulice, J. L., & Kadoch, C. (2016). Composition and Function of Mammalian SWI/SNF Chromatin Remodeling Complexes in Human Disease. *Cold Spring Harbor Symposia on Quantitative Biology*, *81*, 53–60. <https://doi.org/10.1101/sqb.2016.81.031021>
- Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A., & Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA–genome interactions. *Genes & Development*, *30*(2), 191–207. <https://doi.org/10.1101/gad.272187.115>
- Rahmanto, Y. S., Jung, J.-G., Wu, R.-C., Kobayashi, Y., Heaphy, C. M., Meeker, A. K., Wang, T.-L., & Shih, I.-M. (2016). Inactivating ARID1A Tumor Suppressor Enhances TERT Transcription and Maintains Telomere Length in Cancer Cells. *Journal of Biological Chemistry*, *291*(18), 9690–9699. <https://doi.org/10.1074/jbc.M115.707612>
- Reisman, D. N., Sciarrotta, J., Bouldin, T. W., Weissman, B. E., & Funkhouser, W. K. (2005). The Expression of the SWI/SNF ATPase Subunits BRG1 and BRM in Normal Human Tissues. *Applied Immunohistochemistry & Molecular Morphology*, *13*(1), 66–74. <https://doi.org/10.1097/00129039-200503000-00011>
- Rimsza, L. M., LeBlanc, M. L., Unger, J. M., Miller, T. P., Grogan, T. M., Persky, D. O., Martel, R. R., Sabalos, C. M., Seligmann, B., Braziel, R. M., Campo, E., Rosenwald, A., Connors, J. M., Sehn, L. H., Johnson, N., & Gascoyne, R. D. (2008). Gene expression predicts overall survival in paraffin-embedded tissues of diffuse large B-cell lymphoma treated with R-

- CHOP. *Blood*, 112(8), 3425–3433. <https://doi.org/10.1182/blood-2008-02-137372>
- Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, 81(1), 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>
- Roberts, K. G. (2017). The biology of Philadelphia chromosome-like ALL. *Best Practice & Research Clinical Haematology*, 30(3), 212–221. <https://doi.org/10.1016/j.beha.2017.07.003>
- Roberts, K. G. (2018). Genetics and prognosis of ALL in children vs adults. *Hematology*, 2018(1), 137–145. <https://doi.org/10.1182/asheducation-2018.1.137>
- Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, 6(8), 597–610. <https://doi.org/10.1038/nrg1655>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rodríguez-Malavé, N. I., Fernando, T. R., Patel, P. C., Contreras, J. R., Palanichamy, J. K., Tran, T. M., Anguiano, J., Davoren, M. J., Alberti, M. O., Pioli, K. T., Sandoval, S., Crooks, G. M., & Rao, D. S. (2015). BALR-6 regulates cell growth and cell survival in B-lymphoblastic leukemia. *Molecular Cancer*, 14(1), 214. <https://doi.org/10.1186/s12943-015-0485-z>
- Rudant, J., Orsi, L., Menegaux, F., Petit, A., Baruchel, A., Bertrand, Y., Lambilliotte, A., Robert, A., Michel, G., Margueritte, G., Tandonnet, J., Mechinaud, F., Bordigoni, P., Hemon, D., & Clavel, J. (2010). Childhood Acute Leukemia, Early Common Infections, and Allergy: The ESCALE Study. *American Journal of Epidemiology*, 172(9), 1015–1027. <https://doi.org/10.1093/aje/kwq233>
- Safavi, S., & Paulsson, K. (2017). Near-haploid and low-hypodiploid acute lymphoblastic leukemia: two distinct subtypes with consistently poor prognosis. *Blood*, 129(4), 420–423. <https://doi.org/10.1182/blood-2016-10-743765>
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell*, 146(3), 353–358. <https://doi.org/10.1016/j.cell.2011.07.014>
- Sandhya, S., Maulik, A., Giri, M., & Singh, M. (2018). Domain architecture of BAF250a reveals the ARID and ARM-repeat domains with implication in

- function and assembly of the BAF remodeling complex. *PLOS ONE*, *13*(10), e0205267. <https://doi.org/10.1371/journal.pone.0205267>
- Schäfer, D., Olsen, M., Lähnemann, D., Stanulla, M., Slany, R., Schmiegelow, K., Borkhardt, A., & Fischer, U. (2018). Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood*, *131*(7), 821–826. <https://doi.org/10.1182/blood-2017-09-808402>
- Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., & Proudfoot, N. J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, *65*(1), 25–38. <https://doi.org/10.1016/j.molcel.2016.11.029>
- Schultz, I., & Keita, Å. V. (2020). The Intestinal Barrier and Current Techniques for the Assessment of Gut Permeability. *Cells*, *9*(8), 1909. <https://doi.org/10.3390/cells9081909>
- Schwind, S., Marcucci, G., Kohlschmidt, J., Radmacher, M. D., Mrózek, K., Maharry, K., Becker, H., Metzeler, K. H., Whitman, S. P., Wu, Y.-Z., Powell, B. L., Baer, M. R., Kolitz, J. E., Carroll, A. J., Larson, R. A., Caligiuri, M. A., & Bloomfield, C. D. (2011). Low expression of MN1 associates with better treatment response in older patients with de novo cytogenetically normal acute myeloid leukemia. *Blood*, *118*(15), 4188–4198. <https://doi.org/10.1182/blood-2011-06-357764>
- Shen, H., & Laird, P. W. (2013). Interplay between the Cancer Genome and Epigenome. *Cell*, *153*(1), 38–55. <https://doi.org/10.1016/j.cell.2013.03.008>
- Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., Kapoor, P., Ju, Z., Mo, Q., Shih, I.-M., Uray, I. P., Wu, X., Brown, P. H., Shen, X., Mills, G. B., & Peng, G. (2015). ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discovery*, *5*(7), 752–767. <https://doi.org/10.1158/2159-8290.CD-14-0849>
- Shi, H., Li, C., Feng, W., Yue, J., Song, J., Peng, A., & Wang, H. (2020). BCL11A Is Oncogenic and Predicts Poor Outcomes in Natural Killer/T-Cell Lymphoma. *Frontiers in Pharmacology*, *11*. <https://doi.org/10.3389/fphar.2020.00820>
- Shimo, T., Nakanishi, T., Nishida, T., Asano, M., Sasaki, A., Kanyama, M., Kuboki, T., Matsumura, T., & Takigawa, M. (2001). Involvement of CTGF, a Hypertrophic Chondrocyte-Specific Gene Product, in Tumor Angiogenesis. *Oncology*, *61*(4), 315–322. <https://doi.org/10.1159/000055339>

- Shojaee, S., Garcia, C., Wu, H., & Muschen, M. (2010). The Tumor Suppressor PTEN Is Required to Prevent Cellular Senescence and Cell Cycle Arrest In B Cell Lineage and Chronic Myeloid Leukemia. *Blood*, *116*(21), 513–513. <https://doi.org/10.1182/blood.V116.21.513.513>
- Shu, X. O., Potter, J. D., Linet, M. S., Severson, R. K., Han, D., Kersey, J. H., Neglia, J. P., Trigg, M. E., & Robison, L. L. (2002). Diagnostic X-rays and ultrasound exposure and risk of childhood acute lymphoblastic leukemia by immunophenotype. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *11*(2), 177–185.
- Shu, X. O., Reaman, G. H., Lampkin, B., Sather, H. N., Pendergrass, T. W., & Robison, L. L. (1994). Association of paternal diagnostic X-ray exposure with risk of infant leukemia. Investigators of the Childrens Cancer Group. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *3*(8), 645–653.
- Singh, A., Modak, S. B., Chaturvedi, M. M., & Purohit, J. S. (2023). SWI/SNF Chromatin Remodelers: Structural, Functional and Mechanistic Implications. *Cell Biochemistry and Biophysics*, *81*(2), 167–187. <https://doi.org/10.1007/s12013-023-01140-5>
- Smith, A. L., Jung, E.-M., Jeon, B. T., & Kim, W.-Y. (2020). Arid1b haploinsufficiency in parvalbumin- or somatostatin-expressing interneurons leads to distinct ASD-like and ID-like behavior. *Scientific Reports*, *10*(1), 7834. <https://doi.org/10.1038/s41598-020-64066-5>
- Spector, L. G., Ross, J. A., Robison, L. L., & Bhatia, S. (2006). Epidemiology and etiology. In *Childhood Leukemias* (pp. 48–66). Cambridge University Press. <https://doi.org/10.1017/CBO9780511471001.004>
- Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology*, *22*(2), 96–118. <https://doi.org/10.1038/s41580-020-00315-9>
- Statello, L., Maugeri, M., Garre, E., Nawaz, M., Wahlgren, J., Papadimitriou, A., Lundqvist, C., Lindfors, L., Collén, A., Sunnerhagen, P., Ragusa, M., Purrello, M., Di Pietro, C., Tigue, N., & Valadi, H. (2018). Identification of RNA-binding proteins in exosomes capable of interacting with different types of RNA: RBP-facilitated transport of RNAs into exosomes. *PLOS ONE*, *13*(4), e0195969. <https://doi.org/10.1371/journal.pone.0195969>

- Steinhilber, D., & Marschalek, R. (2018). How to effectively treat acute leukemia patients bearing MLL-rearrangements? *Biochemical Pharmacology*, *147*, 183–190. <https://doi.org/10.1016/j.bcp.2017.09.007>
- Stephens, M. (2016). False discovery rates: a new deal. *Biostatistics*, kxw041. <https://doi.org/10.1093/biostatistics/kxw041>
- Sun, X., Chuang, J.-C., Kanchwala, M., Wu, L., Celen, C., Li, L., Liang, H., Zhang, S., Maples, T., Nguyen, L. H., Wang, S. C., Signer, R. A. J., Sorouri, M., Nassour, I., Liu, X., Xu, J., Wu, M., Zhao, Y., Kuo, Y.-C., ... Zhu, H. (2016). Suppression of the SWI/SNF Component Arid1a Promotes Mammalian Regeneration. *Cell Stem Cell*, *18*(4), 456–466. <https://doi.org/10.1016/j.stem.2016.03.001>
- Sun, X., Wang, S. C., Wei, Y., Luo, X., Jia, Y., Li, L., Gopal, P., Zhu, M., Nassour, I., Chuang, J.-C., Maples, T., Celen, C., Nguyen, L. H., Wu, L., Fu, S., Li, W., Hui, L., Tian, F., Ji, Y., ... Zhu, H. (2017). Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. *Cancer Cell*, *32*(5), 574–589.e6. <https://doi.org/10.1016/j.ccell.2017.10.007>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Supek, F., Bošnjak, M., Kunca, S., & Muc, S. (2011). Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), 21800. <https://doi.org/10.1371/journal.pone.0021800>
- Takeda, R., Asada, S., Park, S.-J., Yokoyama, A., Becker, H. J., Kanai, A., Visconte, V., Hershberger, C. E., Hayashi, Y., Yonezawa, T., Tamura, M., Fukushima, T., Tanaka, Y., Fukuyama, T., Matsumoto, A., Yamasaki, S., Nakai, K., Yamazaki, S., Inaba, T., ... Kitamura, T. (2020). HHEX promotes myeloid transformation in cooperation with mutant ASXL1. *Blood*. <https://doi.org/10.1182/blood.2019004613>
- Takeshima, H., & Ushijima, T. (2019). Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. *Npj Precision Oncology*, *3*(1), 7. <https://doi.org/10.1038/s41698-019-0079-0>
- Tang, Q., & Hann, S. S. (2018). HOTAIR: An Oncogenic Long Non-Coding RNA in Human Cancer. *Cellular Physiology and Biochemistry*, *47*(3), 893–913. <https://doi.org/10.1159/000490131>

- Tan-Wong, S. M., Dhir, S., & Proudfoot, N. J. (2019). R-Loops Promote Anti-sense Transcription across the Mammalian Genome. *Molecular Cell*, 76(4), 600-616.e6. <https://doi.org/10.1016/j.molcel.2019.10.002>
- Tauchi, H., Tomizawa, D., Eguchi, M., Eguchi-Ishimae, M., Koh, K., Hirayama, M., Miyamura, N., Kinukawa, N., Hayashi, Y., Horibe, K., & Ishii, E. (2008). Clinical features and outcome of MLL gene rearranged acute lymphoblastic leukemia in infants with additional chromosomal abnormalities other than 11q23 translocation. *Leukemia Research*, 32(10), 1523–1529. <https://doi.org/10.1016/j.leukres.2008.03.018>
- Tebbi, C. K. (2021). Etiology of Acute Leukemia: A Review. *Cancers*, 13(9), 2256. <https://doi.org/10.3390/cancers13092256>
- Terwilliger, T., & Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer Journal*, 7(6), e577–e577. <https://doi.org/10.1038/bcj.2017.53>
- Tirado, C. A., Shabsovich, D., Yeh, L., Pullarkat, S. T., Yang, L., Kallen, M., & Rao, N. (2015). A (1;19) translocation involving TCF3-PBX1 fusion within the context of a hyperdiploid karyotype in adult B-ALL: a case report and review of the literature. *Biomarker Research*, 3(1), 4. <https://doi.org/10.1186/s40364-015-0029-0>
- Tirindelli, R., Dibattista, M., Pifferi, S., & Menini, A. (2009). From Pheromones to Behavior. *Physiological Reviews*, 89(3), 921–956. <https://doi.org/10.1152/physrev.00037.2008>
- Tolstorukov, M. Y., Sansam, C. G., Lu, P., Koellhoffer, E. C., Helming, K. C., Alver, B. H., Tillman, E. J., Evans, J. A., Wilson, B. G., Park, P. J., & Roberts, C. W. M. (2013). Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proceedings of the National Academy of Sciences*, 110(25), 10165–10170. <https://doi.org/10.1073/pnas.1302209110>
- Topatana, W., Juengpanich, S., Li, S., Cao, J., Hu, J., Lee, J., Suliyanto, K., Ma, D., Zhang, B., Chen, M., & Cai, X. (2020). Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *Journal of Hematology & Oncology*, 13(1), 118. <https://doi.org/10.1186/s13045-020-00956-5>
- Trizzino, M., Barbieri, E., Petracovici, A., Wu, S., Welsh, S. A., Owens, T. A., Licciulli, S., Zhang, R., & Gardini, A. (2018). The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II. *Cell Reports*, 23(13), 3933–3945. <https://doi.org/10.1016/j.celrep.2018.05.097>

- Tu, Z., & Zheng, Y. (2022). Role of ATP-dependent chromatin remodelers in hematopoietic stem and progenitor cell maintenance. *Current Opinion in Hematology*, 29(4), 174–180. <https://doi.org/10.1097/MOH.0000000000000710>
- Uszczyńska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., & Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nature Reviews Genetics*, 19(9), 535–548. <https://doi.org/10.1038/s41576-018-0017-y>
- van der Sluijs, P. J., Jansen, S., Vergano, S. A., Adachi-Fukuda, M., Alanay, Y., AlKindy, A., Baban, A., Bayat, A., Beck-Wödl, S., Berry, K., Bijlsma, E. K., Bok, L. A., Brouwer, A. F. J., van der Burgt, I., Campeau, P. M., Canham, N., Chrzanowska, K., Chu, Y. W. Y., Chung, B. H. Y., ... Santen, G. W. E. (2019). The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin–Siris syndrome. *Genetics in Medicine*, 21(6), 1295–1307. <https://doi.org/10.1038/s41436-018-0330-z>
- van Kalsbeek, R. J., Hudson, M. M., Mulder, R. L., Ehrhardt, M., Green, D. M., Mulrooney, D. A., Hakkert, J., den Hartogh, J., Nijenhuis, A., van Santen, H. M., Schouten-van Meeteren, A. Y. N., van Tinteren, H., Verbruggen, L. C., Conklin, H. M., Jacola, L. M., Webster, R. T., Partanen, M., Kollen, W. J. W., Grootenhuis, M. A., ... van Tinteren, H. (2023). A joint international consensus statement for measuring quality of survival for patients with childhood cancer. *Nature Medicine*, 29(6), 1340–1348. <https://doi.org/10.1038/s41591-023-02339-y>
- Van Rechem, C., Boulay, G., & Leprince, D. (2009). HIC1 interacts with a specific subunit of SWI/SNF complexes, ARID1A/BAF250A. *Biochemical and Biophysical Research Communications*, 385(4), 586–590. <https://doi.org/10.1016/j.bbrc.2009.05.115>
- Vendramin, R., Litchfield, K., & Swanton, C. (2021). Cancer evolution: Darwin and beyond. *The EMBO Journal*, 40(18). <https://doi.org/10.15252/emj.2021108389>
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M. U., Ohgi, K. A., Glass, C. K., Rosenfeld, M. G., & Fu, X.-D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, 474(7351), 390–394. <https://doi.org/10.1038/nature10006>
- Wang, F.-Y., Gu, Z.-Y., & Gao, C.-J. (2020). Emerging role of long non-coding RNAs in normal and malignant hematopoiesis. *Chinese Medical Journal*, 133(4), 462–473. <https://doi.org/10.1097/CM9.0000000000000624>

- Wang, J., Ma, R., Ma, W., Chen, J., Yang, J., Xi, Y., & Cui, Q. (2016). LncDisease: a sequence based bioinformatics tool for predicting lncRNA-disease associations. *Nucleic Acids Research*, *44*(9), e90–e90. <https://doi.org/10.1093/nar/gkw093>
- Wang, L., Fu, D., Qiu, Y., Xing, X., Xu, F., Han, C., Xu, X., Wei, Z., Zhang, Z., Ge, J., Cheng, W., & Xie, H.-L. (2014). Genome-wide screening and identification of long noncoding RNAs and their interaction with protein coding RNAs in bladder urothelial cell carcinoma. *Cancer Letters*, *349*(1), 77–86. <https://doi.org/10.1016/j.canlet.2014.03.033>
- Wang, L.-H., Wu, C.-F., Rajasekaran, N., & Shin, Y. K. (2018). Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. *Cellular Physiology and Biochemistry*, *51*(6), 2647–2693. <https://doi.org/10.1159/000495956>
- Wang, X., Haswell, J. R., & Roberts, C. W. M. (2014). Molecular Pathways: SWI/SNF (BAF) Complexes Are Frequently Mutated in Cancer—Mechanisms and Potential Therapeutic Insights. *Clinical Cancer Research*, *20*(1), 21–27. <https://doi.org/10.1158/1078-0432.CCR-13-0280>
- Wang, X., Nagl, N. G., Flowers, S., Zweitzig, D., Dallas, P. B., & Moran, E. (2004). Expression of p270(ARID1A), a component of human SWI/SNF complexes, in human tumors. *International Journal of Cancer*, *112*(4), 636–642. <https://doi.org/10.1002/ijc.20450>
- Wang, X., Wang, S., Troisi, E. C., Howard, T. P., Haswell, J. R., Wolf, B. K., Hawk, W. H., Ramos, P., Oberlick, E. M., Tzvetkov, E. P., Ross, A., Vazquez, F., Hahn, W. C., Park, P. J., & Roberts, C. W. M. (2019). BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nature Communications*, *10*(1), 1881. <https://doi.org/10.1038/s41467-019-09891-7>
- Wang, Z., Chen, K., Jia, Y., Chuang, J.-C., Sun, X., Lin, Y.-H., Celen, C., Li, L., Huang, F., Liu, X., Castrillon, D. H., Wang, T., & Zhu, H. (2020). Dual ARID1A/ARID1B loss leads to rapid carcinogenesis and disruptive redistribution of BAF complexes. *Nature Cancer*, *1*(9), 909–922. <https://doi.org/10.1038/s43018-020-00109-0>
- Wang, Z., Wang, P., Li, Y., Peng, H., Zhu, Y., Mohandas, N., & Liu, J. (2021). Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. *Signal Transduction and Targeted Therapy*, *6*(1), 24. <https://doi.org/10.1038/s41392-020-00422-1>
- Wanior, M., Krämer, A., Knapp, S., & Joerger, A. C. (2021). Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy.

- Oncogene*, 40(21), 3637–3654. <https://doi.org/10.1038/s41388-021-01781-x>
- Wapinski, O., & Chang, H. Y. (2011). Long noncoding RNAs and human disease. *Trends in Cell Biology*, 21(6), 354–361. <https://doi.org/10.1016/j.tcb.2011.04.001>
- Watabe, K. (2015). Roles of lncRNA in breast cancer. *Frontiers in Bioscience*, 7(1), 427. <https://doi.org/10.2741/s427>
- Wells, J. E., Howlett, M., Cheung, L. C., & Kees, U. R. (2015). The role of CCN family genes in haematological malignancies. *Journal of Cell Communication and Signaling*, 9(3), 267–278. <https://doi.org/10.1007/s12079-015-0296-4>
- West, J., Schenck, R. O., Gatenbee, C., Robertson-Tessi, M., & Anderson, A. R. A. (2021). Normal tissue architecture determines the evolutionary course of cancer. *Nature Communications*, 12(1), 2060. <https://doi.org/10.1038/s41467-021-22123-1>
- Wiles, E. T., & Selker, E. U. (2017). H3K27 methylation: a promiscuous repressive chromatin mark. *Current Opinion in Genetics & Development*, 43, 31–37. <https://doi.org/10.1016/j.gde.2016.11.001>
- Witkowski, L., Carrot-Zhang, J., Albrecht, S., Fahiminiya, S., Hamel, N., Tomiak, E., Grynspan, D., Saloustros, E., Nadaf, J., Rivera, B., Gilpin, C., Castellsagué, E., Silva-Smith, R., Plourde, F., Wu, M., Saskin, A., Arseneault, M., Karabakhtsian, R. G., Reilly, E. A., ... Foulkes, W. D. (2014). Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nature Genetics*, 46(5), 438–443. <https://doi.org/10.1038/ng.2931>
- Wu, J. N., & Roberts, C. W. M. (2013). *ARID1A* Mutations in Cancer: Another Epigenetic Tumor Suppressor? *Cancer Discovery*, 3(1), 35–43. <https://doi.org/10.1158/2159-8290.CD-12-0361>
- Xiang, J.-F., Yin, Q.-F., Chen, T., Zhang, Y., Zhang, X.-O., Wu, Z., Zhang, S., Wang, H.-B., Ge, J., Lu, X., Yang, L., & Chen, L.-L. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Research*, 24(5), 513–531. <https://doi.org/10.1038/cr.2014.35>
- Xu, F., Flowers, S., & Moran, E. (2012). Essential Role of ARID2 Protein-containing SWI/SNF Complex in Tissue-specific Gene Expression. *Journal of Biological Chemistry*, 287(7), 5033–5041. <https://doi.org/10.1074/jbc.M111.279968>

- Yang, F., Tuxhorn, J. A., Ressler, S. J., McAlhany, S. J., Dang, T. D., & Rowley, D. R. (2005). Stromal Expression of Connective Tissue Growth Factor Promotes Angiogenesis and Prostate Cancer Tumorigenesis. *Cancer Research*, 65(19), 8887–8895. <https://doi.org/10.1158/0008-5472.CAN-05-1702>
- Yang, M., Vesterlund, M., Siavelis, I., Moura-Castro, L. H., Castor, A., Fioretos, T., Jafari, R., Lilljebjörn, H., Odom, D. T., Olsson, L., Ravi, N., Woodward, E. L., Harewood, L., Lehtiö, J., & Paulsson, K. (2019). Proteogenomics and Hi-C reveal transcriptional dysregulation in high hyperdiploid childhood acute lymphoblastic leukemia. *Nature Communications* 2019 10:1, 10(1), 1–15. <https://doi.org/10.1038/s41467-019-09469-3>
- Yao, Q., Chen, Y., & Zhou, X. (2019). The roles of microRNAs in epigenetic regulation. *Current Opinion in Chemical Biology*, 51, 11–17. <https://doi.org/10.1016/j.cbpa.2019.01.024>
- Yap, K., Mukhina, S., Zhang, G., Tan, J. S. C., Ong, H. S., & Makeyev, E. V. (2018). A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Molecular Cell*, 72(3), 525–540.e13. <https://doi.org/10.1016/j.molcel.2018.08.041>
- Yousefi, H., Maheronnaghsh, M., Molaei, F., Mashouri, L., Reza Aref, A., Momeny, M., & Alahari, S. K. (2020). Long noncoding RNAs and exosomal lncRNAs: classification, and mechanisms in breast cancer metastasis and drug resistance. *Oncogene*, 39(5), 953–974. <https://doi.org/10.1038/s41388-019-1040-y>
- Yu, S.-L., Zhang, H., Ho, B.-C., Yu, C.-H., Chang, C.-C., Hsu, Y.-C., Ni, Y.-L., Lin, K.-H., Jou, S.-T., Lu, M.-Y., Chen, S.-H., Wu, K.-H., Wang, S.-C., Chang, H.-H., Pui, C.-H., Yang, J. J., Zhang, J., Lin, D.-T., Lin, S.-W., ... Yang, Y.-L. (2020). FPGS relapse-specific mutations in relapsed childhood acute lymphoblastic leukemia. *Scientific Reports*, 10(1), 12074. <https://doi.org/10.1038/s41598-020-69059-y>
- Zhang, X., Azhar, G., Zhong, Y., & Wei, J. Y. (2006). Zipzap/p200 is a novel zinc finger protein contributing to cardiac gene regulation. *Biochemical and Biophysical Research Communications*, 346(3), 794–801. <https://doi.org/10.1016/j.bbrc.2006.05.211>
- Zhang, Y., Xu, Y., Feng, L., Li, F., Sun, Z., Wu, T., Shi, X., Li, J., & Li, X. (2016). Comprehensive characterization of lncRNA-mRNA related ceRNA network across 12 major cancers. *Oncotarget*, 7(39), 64148. <https://doi.org/10.18632/ONCOTARGET.11637>

- Zhao, S., Wu, W., Jiang, Z., Tang, F., Ding, L., Xu, W., & Ruan, L. (2022). Roles of ARID1A variations in colorectal cancer: a collaborative review. *Molecular Medicine*, 28(1), 42. <https://doi.org/10.1186/s10020-022-00469-6>
- Zheng, R., Lin, S., Guan, L., Yuan, H., Liu, K., Liu, C., Ye, W., Liao, Y., Jia, J., & Zhang, R. (2018). Long non-coding RNA XIST inhibited breast cancer cell growth, migration, and invasion via miR-155/CDX1 axis. *Biochemical and Biophysical Research Communications*, 498(4), 1002–1008. <https://doi.org/10.1016/j.bbrc.2018.03.104>
- Zhou, Y., Jiang, L., & Rui, L. (2009). Identification of MUP1 as a Regulator for Glucose and Lipid Metabolism in Mice. *Journal of Biological Chemistry*, 284(17), 11152–11159. <https://doi.org/10.1074/jbc.M900754200>
- Zhou, Y., & Rui, L. (2010). *Major Urinary Protein Regulation of Chemical Communication and Nutrient Metabolism* (pp. 151–163). [https://doi.org/10.1016/S0083-6729\(10\)83006-7](https://doi.org/10.1016/S0083-6729(10)83006-7)
- Zhu, K., Liu, Q., Zhou, Y., Tao, C., Zhao, Z., Sun, J., & Xu, H. (2015). Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. *BMC Genomics*, 16(S7), S8. <https://doi.org/10.1186/1471-2164-16-S7-S8>

# Anexo

# Anexo

## Financiación

El desarrollo de la tesis doctoral ha sido financiado por un contrato predoctoral otorgado por la Fundación Benéfica Anticáncer Santa Cándida y San Francisco Javier.

## Publicaciones

### Artículos Originales

1. Peinado P\*, Andrades A\*, Cuadros M\*, Rodríguez MI\*, Coira IF, **García DJ**, Benítez-Cantos MS, Cano C, Rufino-Palomares E, Zarzuela E, Muñoz J, Loidi C, Saiz M, Medina PP. “Multi-omic alterations of the SWI/SNF complex define a clinical subgroup in lung adenocarcinoma”. *Clinical Epigenetics*. 2022 Mar 17;14(1):42. doi: [10.1186/s13148-022-01261-3](https://doi.org/10.1186/s13148-022-01261-3).
2. Cuadros M\*, **García DJ\***, Andrades A\*, Arenas AM, Coira IF, Baliñas-Gavira C, Peinado P, Rodríguez MI, Álvarez-Pérez JC, Ruiz-Cabello F, Camós M, Jiménez-Velasco A, Medina PP. “LncRNA-mRNA Co-Expression Analysis Identifies AL133346.1/CCN2 as Biomarkers in Pediatric B-Cell Acute Lymphoblastic Leukemia”. *Cancers (Basel)*. 2020 Dec 17;12(12):3803. doi: [10.3390/cancers12123803](https://doi.org/10.3390/cancers12123803).
3. Peinado P\*, Andrades A\*, Cuadros M\*, Rodríguez MI\*, Coira IF, **García DJ**, Álvarez-Pérez JC, Baliñas-Gavira C, Arenas AM, Patiño-Mercáu JR, Sanjuan-Hidalgo J, Romero OA, Montuenga LM, Carretero J, Sanchez-Céspedes M, Medina PP. “Comprehensive Analysis of SWI/SNF Inactivation in Lung Adenocarcinoma Cell Models”. *Cancers (Basel)*. 2020 Dec 10;12(12):3712. doi: [10.3390/cancers12123712](https://doi.org/10.3390/cancers12123712).
4. Arenas AM\*, Cuadros M\*, Andrades A, **García DJ**, Coira IF, Rodríguez MI, Baliñas-Gavira C, Peinado P, Álvarez-Pérez JC, Medina PP.

“LncRNA DLG2-AS1 as a Novel Biomarker in Lung Adenocarcinoma”. *Cancers (Basel)*. 2020 Jul 28;12(8):2080. doi: [10.3390/cancers12082080](https://doi.org/10.3390/cancers12082080).

5. Cuadros M\*, Andrades A\*, Coira IF, Rodríguez MI, Álvarez-Pérez JC, Peinado P, Arenas AM, **García DJ**, Jiménez P, Camós M, Jiménez, Velasco A, Medina PP. “Expression of the long non-coding RNA TCL6 is associated with clinical outcome in pediatric B-cell acute lymphoblastic leukemia”. *Blood Cancer Journal*. 2019 Nov 25;9(12):93. doi: [10.1038/s41408-019-0258-9](https://doi.org/10.1038/s41408-019-0258-9).

## **Revisiones y Capítulos de Libros**

1. Andrades A\*, Peinado P\*, Álvarez-Pérez JC\*, Sanjuán-Hidalgo J, **García DJ**, Arenas AM, Matia-González AM, Medina PP. “SWI/SNF complexes in hematological malignancies: biological implications and therapeutic opportunities”. *Molecular Cancer*. 2023 Feb 21;22(1):39. doi: [10.1186/s12943-023-01736-8](https://doi.org/10.1186/s12943-023-01736-8).
2. Arenas AM\*, Andrades A\*, Patiño-Mercau JR, Sanjuán-Hidalgo J, Cuadros M, **García DJ**, Peinado P, Rodríguez MI, Baliñas-Gavira C, Álvarez-Pérez JC, Medina PP. “Opportunities of miRNAs in cancer therapeutics”. In “MicroRNA in Human Malignancies, 1st Edition”; Negrini M, Calin G, Croce C (Eds); Elsevier, Feb 17, 2022; pp. 153-164. ISBN: [9780128222874](https://doi.org/9780128222874).

## **Patentes**

- 1) Medina PP (50%), Cuadros M (20%), **García DJ** (10%), Andrades A (10%), Rodríguez MI (5%), Álvarez-Pérez JC (5%). “Biomarcadores para el diagnóstico, prognosis, prevención, mejora o alivio en el tratamiento de la leucemia linfoblástica aguda de células B pediátrica”. Número Solicitud: P202031253. PCT Application Number: PCT/ES2021/070899. CIP: C12Q/6809.

## Permisos de derecho de autor

### Figura 1: <https://creativecommons.org/licenses/by/4.0/>

Global incidence, mortality and temporal trends of cancer in children: A joinpoint regression analysis

Junjie Huang,<sup>1</sup> Sze Chai Chan,<sup>1</sup> Chun Ho Ngai,<sup>1</sup> Veeleah Lok,<sup>2</sup> Lin Zhang,<sup>3,4</sup> Don Eliseo Lucero-Prisco, III,<sup>5</sup> Wanghong Xu,<sup>6</sup> Zhi-Jie Zheng,<sup>7</sup> Edmar Elcarte,<sup>8</sup> Mellissa Withers,<sup>9</sup> Martin C. S. Wong,<sup>1,4,7</sup> and NCD Global Health Research Group, Association of Pacific Rim Universities (APRU)

▸ Author information ▸ Article notes ▸ Copyright and License information [PMC Disclaimer](#)

[Copyright](#) © 2022 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

This is an open access article under the terms of the <http://creativecommons.org/licenses/by/4.0/> License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

### Figura 2: <http://creativecommons.org/licenses/by/3.0/>

GLOBAL Leukemia in Children 0-14 Statistics 2018, Incidence and Mortality and Human Development Index (HDI): GLOBOCAN Sources and Methods

Seyedeh Mahdieh Namayandeh,<sup>1</sup> Zaher Khazaei,<sup>2</sup> Moslem Lari Najafi,<sup>3</sup> Elham Goodarzi,<sup>4,\*</sup> and Alireza Moslem<sup>5,\*</sup>

▸ Author information ▸ Article notes ▸ Copyright and License information: [PMC Disclaimer](#)

[PMC Copyright notice](#)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, (<http://creativecommons.org/licenses/by/3.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Figura 3: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## Epigenetic drivers of tumourigenesis and cancer metastasis

Aniruddha Chatterjee<sup>a,b</sup>  , Euan J. Rodger<sup>a,b</sup>, Michael R. Eccles<sup>a,b</sup>

Show more 

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.semcancer.2017.08.004> 

[Get rights and content](#) 

Under a Creative Commons license 

 open access

**Figura 4:** <https://creativecommons.org/licenses/by/4.0/>**The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia**Thomas R. Jackson,<sup>1,†</sup> Rebecca E. Ling,<sup>1,†</sup> and Anindita Roy<sup>1,2,\*</sup>

[▶ Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [PMC Disclaimer](#)

[Copyright](#) © 2021 Jackson, Ling and Roy.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**Figura 5:** <https://creativecommons.org/licenses/by/4.0/>**The Prenatal Origin of Childhood Leukemia: Potential Applications for Epidemiology and Newborn Screening**Erin L. Marcotte,<sup>1,2,\*</sup> Logan G. Spector,<sup>1,2</sup> Daniela P. Mendes-de-Almeida,<sup>1,3,4</sup> and Heather H. Nelson<sup>2,5</sup>

[▶ Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [PMC Disclaimer](#)

[Copyright](#) © 2021 Marcotte, Spector, Mendes-de-Almeida and Nelson.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**Figura 6:** Solicitud de permiso a través de correo electrónico a la revista Haematologica.

Daniel García García

mar 08/08, 11:22

marketing@haematologica.org; office@haematologica.org ✕

 Responder a todos | ▼

Good morning,

My name is Daniel Jesús García García, predoctoral student at the University of Granada. I would like to request permission to use an adapted version of Figure 1 from the article "Pediatric Acute Lymphoblastic Leukemia" (<https://doi.org/10.3324/haematol.2020.247031>) for inclusion in the publication of my doctoral thesis.

Thank you in advance,

Best regards,

Daniel.



Giulia Carlini &lt;g.carlini@haematologica.org&gt;

jue 24/08, 12:12



Dear Daniel,

Thank you for the information, if that is the case, please use this mail as a permission granted to your needs.

To whom it may concern:

we herewith grant permission to download and re-use article, figure/table just for the use cited in your message below free of charge: this publication is intended for no-profit use. Any other use is discouraged and it will remain under copyright laws of the Ferrata Storti Foundation, Pavia, Italy. You are also obliged to correctly quote the original article reference.

Best regards

Giulia

## Figura 7: <https://creativecommons.org/licenses/by/4.0/>.

### Exploiting vulnerabilities of SWI/SNF chromatin remodelling complexes for cancer therapy

Marek Wanior,<sup>1,2</sup> Andreas Krämer,<sup>1,2,3</sup> Stefan Knapp,<sup>1,2,3,4</sup> and Andreas C. Joerge,<sup>1,2,4</sup>

► Author information ► Article notes ► Copyright and License information ► PMC Disclaimer

Copyright © The Author(s) 2021

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## Figura 8: <https://creativecommons.org/licenses/by/4.0/>

### Aberrant LncRNA Expression in Leukemia

Jie Gao, Fujue Wang, Pengqiang Wu, Yingying Chen, and Yongqian Jia<sup>✉</sup>

► Author information ► Article notes ► Copyright and License information ► PMC Disclaimer

Copyright © The author(s)

This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.