



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

**Caracterización molecular e impacto clínico de células madre  
leucémicas en pacientes con leucemia mieloide crónica.**

Tesis para optar por el título de Doctor de la Universidad de Buenos Aires

Área Química biológica

Lic. María Sol Ruiz

**Director de tesis:** Dr. Michele Bianchini

**Consejera de estudios:** Dra. Elba Vásquez

**Lugar de trabajo:** Centro de investigaciones oncológicas - Fundación Cáncer (CIO-FUCA)

**Buenos Aires, 2019**

**Fecha de defensa:** 2/7/2019



## **Caracterización molecular e impacto clínico de células madre leucémicas en pacientes con leucemia mieloide crónica.**

### **Resumen**

La leucemia mieloide crónica (LMC) se caracteriza por la activación constitutiva de la quinasa BCR-ABL1. A partir de la introducción del tratamiento con inhibidores de tirosina-quinasa (ITK) específicos, la sobrevida global ha mejorado significativamente. Sin embargo, aún luego de muchos años de tratamiento, los pacientes muestran evidencias de enfermedad residual por la persistencia de una población de células madre leucémicas (LSC). En este trabajo nos propusimos responder dos preguntas acerca de esta población:

1) ¿Es de utilidad clínica la cuantificación de las LSC en los pacientes con LMC a lo largo del tratamiento?

2) ¿Qué características biológicas diferencian a las LSC de sus contrapartes normales, las células madre hematopoyéticas (HSC)?

Respecto a la cuantificación de la población de LSC, observamos que los niveles de carga leucémica inicial eran variables entre pacientes, y disminuían en la fracción de células primitivas y progenitoras a lo largo del tratamiento con ITK. A los seis meses de tratamiento, la carga leucémica en la fracción primitiva fue menor en los pacientes clasificados como óptimos respondedores al tratamiento con ITK, vs. aquellos que presentaron signos de advertencia o falla. Esta metodología, basada en la detección del ARNm de *BCR-ABL1* en colonias provenientes de un cultivo de largo término (6 semanas), no fue lo suficientemente sensible para pacientes en RM profunda. Para resolver esta limitación, desarrollamos una metodología capaz de detectar las LSC a nivel de ADN genómico mediante la amplificación del punto de fusión paciente-específico, permitiendo la independización de los niveles de expresión de *BCR-ABL1*. Este ensayo podría tener relevancia para la clínica, ya que será utilizado para evaluar la carga leucémica en el contexto del primer ensayo clínico de discontinuación del tratamiento con ITK en Argentina.

Respecto a la caracterización molecular de las LSC y HSC, evaluamos los niveles globales de microARNs en dichas fracciones provenientes de pacientes con LMC al diagnóstico, en comparación con la fracción primitiva y progenitora de dadores sanos, mediante la tecnología de secuenciación masiva. Describimos conjuntos de microARNs diferencialmente expresados, y realizamos la validación de un subgrupo de ellos mediante RT-qPCR en una nueva cohorte de pacientes y dadores sanos. Observamos que la correlación entre ambas técnicas fue baja, y discutimos las limitaciones que pudieron estar implicadas en dicho resultado. El análisis funcional *in silico* de los microARNs validados reveló una asociación significativa con vías relacionadas a distintos procesos metabólicos. Finalmente, realizamos una búsqueda de microARNs *novel* a partir de los resultados de la secuenciación masiva, detectando la presencia de un microARN de biogénesis no canónica, derivado del clivaje del ARN pequeño nucleolar SCARNA15.

**Palabras clave:** leucemia mieloide crónica, células madre hematopoyéticas, inhibidores de tirosina-quinasa, microARNs, células madre leucémicas.



## **Molecular characterization and clinical impact of leukemic stem cells in chronic myeloid leukemia patients**

### **Abstract**

Chronic myeloid leukemia (CML) is characterized by constitutive activation of BCR-ABL1 kinase. Since the introduction of treatment with specific tyrosine-kinase inhibitors (TKI), overall survival has improved significantly. However, patients on long-term therapy show signs of residual disease burden due to the presence of leukemic stem cells (LSC). In this thesis, we aimed to answer the following questions regarding LSC:

- 1) Is it of clinical utility to quantify the LSC burden in CML patients receiving TKI treatment?
- 2) What are the biological differences between LSC and their normal counterparts, hematopoietic stem cells (HSC)?

Regarding the quantification of LSC, we detected variable levels of leukemic burden in the primitive fraction at diagnosis; we also observed that leukemic burden was gradually reduced along TKI treatment in the primitive and progenitor fraction. At six months of treatment, LSC burden was lower in patients that showed an optimal response to TKI treatment vs. patients that showed signs of warning or failure. For patients showing deep molecular responses in peripheral blood, this methodology, based on the detection of *BCR-ABL1* mRNA-positive colonies in long-term (6-weeks) culture initiating cells (LTC-IC) assays, was not sensitive enough. Therefore, we developed a method that enabled us to detect LSC at the DNA level, independently of *BCR-ABL1* expression levels. In order to design patient-specific primers, we amplified and sequenced the genomic *BCR-ABL1* breakpoint; primers were used to evaluate single colonies from a 6-week LTC-IC assay, by direct nested-PCR. In the context of an ongoing TKI-stop trial in Argentina, this method could be of clinical utility to assess LSC burden before discontinuation.

Regarding characterization of LSC and HSC from CML patients, we analyzed global microRNA expression by massive parallel sequencing. We also characterized primitive and progenitor fractions from healthy donors. After differential expression analysis, we validated a set of microRNAs by RT-qPCR, in a new cohort of patients and healthy donors. We observed a low correlation between both techniques, and we discussed possible limitations related to this result. *In silico* analysis of validated microRNAs revealed a significant association with metabolic pathways. Finally, we searched for novel microRNAs, and found a microRNA of non-canonical biogenesis, derived from the cleavage of a longer molecule, previously characterized as SCARNA15.

**Key words:** chronic myeloid leukemia, hematopoietic stem cells, tyrosine-kinase inhibitors, microRNAs, leukemic stem cells



## PUBLICACIONES

A partir del trabajo realizado durante el doctorado se publicaron los siguientes artículos:

"Evaluation of the primitive fraction by functional in vitro assays at the RNA and DNA level represents a novel tool for complementing molecular monitoring in chronic myeloid leukemia". Autores: Ruiz MS, Sanchez MB, Gutiérrez L, Koile D, Yankilevich P, Mosqueira C, Cranco S, Custidiano MR, Freitas J, Foncuberta C, Moiraghi B, Pavlovsky C, Pérez MA, Ventriglia V, Sanchez Ávalos J, Mordoh J, Larripa I, Bianchini M. *Oncotarget* Vol. 9 (No. 29), 20255-20264 (2018). doi: 10.18632/oncotarget.24749

"Recomendaciones metodológicas para el monitoreo molecular de pacientes con leucemia mieloide crónica por PCR cuantitativa en tiempo real del BCR-ABL1". Autores: Larripa I, Ruiz MS, Gutierrez M, Bianchini M. *Revista Medicina Buenos Aires*, 77 (1), 61-72 (2017)

"Standardization of molecular monitoring for chronic myeloid leukemia in Latin America using locally produced secondary cellular calibrators", Autores: Ruiz MS, Medina M, Tapia I, Mordoh J, Cross NCP, Larripa Irene, Bianchini M. *Leukemia* 30, 2258-2260 (2016) doi:10.1038/leu.2016.197.



## AGRADECIMIENTOS

Creo que esta sección es una de las más importantes de la tesis. El camino por el que uno transcurre durante el doctorado es largo, con dificultades, con momentos gratificantes y momentos frustrantes....se necesita mucha paciencia, sin duda, y sin la ayuda y apoyo de otras personas, sería imposible. O al menos así fue para mí. Así que aquí va una extensa lista de agradecimientos, aunque seguro me estaré olvidando de alguien y pido disculpas de antemano por eso.

Como egresada de la carrera de grado, y realizando el doctorado en la Universidad de Buenos Aires, estaré siempre agradecida con todas las personas e instituciones que permiten que se mantenga en el tiempo como una universidad pública, laica y gratuita, brindando una educación de calidad, un espacio de reflexión y de discusión, un ejemplo para el resto del mundo de que existen alternativas al modelo de universidad empresarial que las grandes corporaciones intentan instalar.

A las instituciones que financiaron distintas etapas del proyecto, de forma directa o indirecta, a través del apoyo a nuestro laboratorio y el otorgamiento de becas: CONICET, ANPCyT, INC, Fundación Sales, FUCA, Fundación Fiorini, Novartis y Fundación Mosoteguy.

Quisiera reconocer a todas las personas que formalmente contribuyeron en distintas etapas del proyecto: Simone Bonecker y Carolina Furtado del INCA (Brasil) colaboraron en la secuenciación del miRnoma; Daniel Koile y Patricio Yankilevich colaboraron en el análisis bioinformático de los datos provenientes de la secuenciación masiva; María Belén Sánchez, Sofía Baffa Trasci, Leandro Gutiérrez y Celeste Mosqueira colaboraron en la amplificación y secuenciación del punto de ruptura BCR-ABL1 a nivel genómico.

A los médicos del Instituto Alexander Fleming, Hospital Posadas, Hospital Ramos Mejía, Hospital de Agudos Prof. Rossi, y Fundaleu, que contribuyeron con muestras de pacientes y datos clínicos, y al servicio de Hemoterapia del Instituto Alexander Fleming, que contribuyó con muestras de dadores sanos. A los pacientes que aceptaron participar del estudio.

Muchas gracias al Dr. Mordoh por haberme abierto las puertas de su laboratorio, y por su comprensión, demostrando siempre un gran compromiso y dedicación con el trabajo y con sus pacientes. De la misma forma, le agradezco a Michele por haberme introducido en el mundo de la LMC, gracias por haber confiado en mí y hacerme partícipe de los múltiples proyectos, por respetar mis tiempos y escuchar mis planteos. ¡Gracias!

Muchas gracias a Irene Larripa, por su apoyo constante, por todos los estudios citogenéticos, por el trabajo conjunto en el proyecto de calibradores, por las charlas y discusiones.

En el laboratorio me encontré con personas increíbles, sin duda todo hubiese sido más difícil sin ellas. Gracias a todas y todos por las muchísimas charlas, debates, salidas, orejas cuando las cosas no salían, ayuda y colaboración constante. Porque las tesis son en realidad el resultado de un trabajo colectivo en el que participaron, en distinta medida, muchísimas personas. A las y los que se fueron, todos contribuyeron en algún momento y dejaron su huella: Pau R, Ivi, Gaby, Yami, Vero, Betu, Michelle, Chervo, Juan, Chofi, Ani. A

los que siguen y mantienen vivo el laboratorio, alegrando el día a día: Paulita, Belu, Estefi, Kike, Holly, Masiel, Sofi, Maru, Julián y Marian. A Estrella y Marcela por sus consejos, siempre con la mejor onda y predisposición. A Dani, Caro y Sil, porque con su trabajo nos ayudan muchísimo! Y siempre con una sonrisa. Un agradecimiento especial a Belu, ya que sin su ayuda esta tesis no estaría terminada al día de hoy. Tuve mucha suerte de que te hayas cruzado en mi camino, ¡gracias por todo!

A Flor, Anto e Ivi, compañeras de docencia, con las que formamos un gran equipo. Gracias por la mejor predisposición para trabajar, y para charlar! Voy a extrañar nuestros debates sobre el sistema científico y nuestro futuro laboral.

A las chicas de FIL: Sofi, Sole, Anto, Flor, Lau, por toda su ayuda cada vez que necesité hacer algún experimento allá.

A Guillermo Piazza por su asistencia técnica en el sorter, especialmente por su paciencia con mis consideraciones (obsesivas) para poder obtener fracciones puras!

A mis amigos de la vida, del taller de tango, a Recu, Cari y Soni, por acompañarme y darme apoyo durante todos estos años.

A mis compañeros de apoyo escolar, Colectivo Sur e IPPM, todas personas con un gran compromiso social, a quienes admiro mucho. Gracias por haberme mostrado que existen otras miradas posibles de la realidad.

Por último, pero no menos importante, agradecer al sostén fundamental de mi vida, mi familia. Gracias a mis padres por haberme ayudado, incentivado, siempre, a hacer lo que me gustaba, y a hacerlo con compromiso, dedicación y honestidad. Por darme todas las oportunidades. A mi papá por haber motivado este tema de investigación, sin llegar a saberlo. A mi mamá por estar ahí siempre, especialmente ahora que Maite llegó a nuestras vidas. A mis hermanos Javier y Mariano, por su generosidad, por bancarla, siempre, en las buenas y en las malas. A mi abuelo, a obachan, a Mari y Cami, a mis tíos y primos por su apoyo incondicional. A mi nueva familia, los Hoqui, por todo el apoyo y ayuda que me dieron siempre, especialmente con Maite en este último tiempo. A Maite, por iluminar mi vida todos los días. A Martín, quien me acompañó durante todo este trayecto, gracias por todo el amor, por ser mi sostén, por amar y cuidar a Maite incondicionalmente. Sin tu ayuda esta tesis no hubiese sido posible. Gracias infinitas.

## ABREVIATURAS

<b>ADNg</b>	ADN genómico
<b>ARNm</b>	ARN mensajero
<b>ARNt</b>	ARN de transferencia
<b>CFU</b>	Unidades formadoras de colonias
<b>CMP</b>	Progenitores comunes mieloides
<b>CSC</b>	<i>Cancer stem cell</i>
<b>DS</b>	Donantes sanos
<b>FC</b>	<i>Fold change</i>
<b>G-CSF</b>	Factor estimulante de colonias de granulocito
<b>GMP</b>	Progenitores comprometidos a linaje granulocito-macrófago
<b>HSC</b>	Célula madre hematopoyética
<b>IL-3</b>	Interleuquina-3
<b>IM</b>	Imatinib
<b>IRIS</b>	<i>International Randomized Study of Interferon and STI571</i>
<b>IS</b>	Escala internacional
<b>ITK</b>	Inhibidor de tirosina-quinasa
<b>LMC</b>	Leucemia mieloide crónica
<b>LSC</b>	Célula madre leucémica
<b>LTC</b>	Cultivo de largo término
<b>LTC-IC</b>	Célula iniciadora de cultivo de largo término
<b>MEP</b>	Progenitores comprometidos a linaje megacariocito-eritrocito
<b>MLP</b>	Progenitores multi-linfoides
<b>MO</b>	Médula ósea
<b>MPP</b>	Progenitores multipotentes
<b>NK</b>	<i>Natural killer</i>
<b>Ph</b>	Cromosoma Philadelphia
<b>qPCR</b>	PCR en tiempo real
<b>RISC</b>	<i>RNA-induced silencing complex</i>
<b>RM</b>	Respuesta molecular
<b>RT</b>	Retrotranscripción
<b>SCARNA</b>	ARN pequeño específico de cuerpos de Cajal
<b>SCF</b>	<i>Stem cell factor</i>
<b>snARN</b>	ARN pequeño nuclear
<b>snoARN</b>	ARN pequeño nucleolar
<b>SP</b>	Sangre periférica
<b>TFR</b>	Remisión libre de tratamiento
<b>UMRD</b>	enfermedad mínima residual indetectable
<b>UTR</b>	<i>Untranslated region</i>



## INDICE

CAPÍTULO I. INTRODUCCIÓN.....	15
I.1. Leucemia mieloide crónica.....	15
I.1.2. Epidemiología.....	15
I.1.3. Características del rearrreglo genómico <i>BCR-ABL1</i> .....	15
I.1.4. Diagnóstico.....	16
I.1.5. Fisiopatología.....	17
I.1.6. Tratamiento.....	18
I.1.7. Desarrollo de inhibidores de <i>BCR-ABL1</i> de segunda generación.....	18
I.1.8. Monitoreo molecular.....	19
I.1.9. Discontinuación.....	22
I.2. Células madre hematopoyéticas.....	22
I.2.1. Modelos para estudiar la funcionalidad de las HSC.....	24
I.2.2. Origen y persistencia del clon leucémico.....	24
I.2.3. Características de las LSC.....	25
I.3. Regulación de la expresión génica mediada por microARNs.....	26
I.3.1. Organización genómica.....	26
I.3.2. Biogénesis de microARNs.....	27
I.3.3. Mecanismos de acción de microARNs.....	28
I.3.4. Participación de microARNs en procesos biológicos.....	28
I.4. Antecedentes de la hipótesis.....	29
HIPÓTESIS.....	30
MATERIALES Y MÉTODOS.....	31
II.1. Muestras de pacientes.....	31
II.2. Aislamiento de la fracción CD34 <sup>+</sup> .....	31
II.3. Ensayo para células progenitoras (CFU).....	31
II.4. Ensayo para células primitivas (LTC-IC).....	33
II.4.1. Cultivo de las líneas M2-10B4 y SI/SI.....	34
II.5. Extracción de ARN y RT-qPCR para la detección de transcritos <i>BCR-ABL1</i> .....	34
II.6. Secuenciación masiva de la fracción de ARNs pequeños.....	35
II.6.1. Citometría de flujo y <i>sorting</i> .....	35
II.6.2. Concentración del ARN proveniente de distintas muestras para la secuenciación masiva del miRnoma.....	36
II.6.3. Construcción de bibliotecas de ARNs pequeños para secuenciación masiva en HiSeq 2500 (Illumina).....	37
II.6.4. Análisis bioinformático de los datos de secuenciación masiva del miRnoma.....	38
II.7. Detección del punto de ruptura <i>BCR-ABL1</i> por secuenciación masiva.....	39
II.7.1. Preparación de muestras para la generación de bibliotecas.....	39
II.7.2. Análisis bioinformático de los datos de secuenciación masiva para la identificación del punto de fusión <i>BCR-ABL1</i> .....	40
II.7.3. Detección de células <i>BCR-ABL1</i> <sup>+</sup> en la fracción primitiva a nivel de ADN genómico.....	40
II.8. Cuantificación de la expresión de microARNs por stem-loop RT-qPCR.....	41
II.9. Análisis estadístico.....	43
RESULTADOS.....	44

<b>Capítulo III. Evaluación de la carga leucémica como biomarcador de respuesta al tratamiento con ITK en la fracción primitiva y progenitora de pacientes con LMC</b>	44
III.1. Introducción	44
III.2. Pacientes incluidos en el estudio y puesta a punto del ensayo	44
III.3. Evaluación de la carga leucémica en pacientes con LMC al diagnóstico y durante el tratamiento con ITK	49
III.4. Comparación de la carga leucémica en pacientes con distinta respuesta clínica al tratamiento con ITK	50
III.5. Detección de LSC basada en la amplificación del punto de fusión <i>BCR-ABL1</i> a nivel de ADN genómico	54
III.5.1. Puesta a punto de la secuenciación del punto de ruptura <i>BCR-ABL1</i> a nivel genómico	54
III.5.2. Detección de la carga leucémica en la fracción primitiva a nivel de ADN genómico	56
<b>Capítulo IV. Caracterización y validación del miRNoma de LSC y HSC de pacientes con LMC al diagnóstico.</b>	66
IV. 1. Introducción	66
IV. 2. Puesta a punto del aislamiento de la fracción de LSC y HSC de pacientes con LMC al diagnóstico	67
IV. 3 Construcción de la biblioteca de ARNs pequeños para secuenciación masiva	73
IV. 4. Resultados de la secuenciación masiva del miRNoma de HSC y LSC.	75
IV. 4.1. Análisis de calidad de los datos	75
IV. 4.2. Análisis bioinformático para la identificación de microARNs	76
IV.5. Validación de los resultados de secuenciación masiva por RT-qPCR.	82
IV.5.1 Puesta a punto de la detección por RT-qPCR multiplex	83
IV.5.2. Selección de pacientes	84
IV.5.3. Evaluación del perfil de citometría de flujo en pacientes con distinta respuesta al tratamiento con ITKs	86
IV.5.4. Cuantificación de los niveles de microARNs por RT-qPCR.	88
IV.5.4.1. Análisis de expresión diferencial	88
IV.5.4.2. Correlación entre secuenciación masiva y RT-qPCR	90
IV.5.5. Análisis multivariado	93
IV.6. Discusión y conclusiones parciales	96
<b>Capítulo V. Análisis <i>in silico</i> de microARNs</b>	102
V.1. Introducción	102
V.2. Análisis de perfiles de expresión	103
V.3. Análisis del novel-3	109
V.4 Análisis de vías asociadas a los microARNs validados por RT-qPCR.	110
V.5. Selección de potenciales genes blanco de los microARNs validados por RT-qPCR.	112
V. 6. Discusión y conclusiones parciales	114
<b>CONCLUSIONES FINALES Y PERSPECTIVAS FUTURAS</b>	119
<b>REFERENCIAS</b>	123
<b>Anexo I. Secuenciación masiva de la fracción de ARNs pequeños</b>	129
<b>Anexo II. Secuencia de primers y sondas</b>	134

## CAPÍTULO I. INTRODUCCIÓN

### I.1. Leucemia mieloide crónica

La Leucemia Mieloide Crónica (LMC) es una enfermedad originada a partir de la transformación oncogénica de precursores hematopoyéticos multipotentes, que se manifiesta como una expansión de células mieloides maduras, principalmente granulocitos, aunque también se observa un incremento de eritroblastos y megacariocitos. Se caracteriza por la presencia del cromosoma Philadelphia (Ph) producto de la translocación recíproca  $t(9;22)(q34;q11)$ , originando el gen de fusión *BCR-ABL1*, el cual codifica una proteína tirosina-quinasa constitutivamente activa<sup>1</sup>. La enfermedad suele presentarse en fase crónica, que puede mantenerse latente por varios años. En ausencia de tratamiento efectivo, la enfermedad progresa, a través de una fase acelerada, hacia una fase de crisis blástica con características similares a una leucemia mieloide aguda (leucocitosis, citopenia, dolor óseo). En estos casos, el pronóstico de los pacientes empeora sustancialmente, siendo el trasplante de células progenitoras hematopoyéticas la única alternativa de curación<sup>2</sup>.

#### I.1.2. Epidemiología

La incidencia de la LMC está reportada como 10 a 15 casos por millón de personas por año, sin una distribución geográfica o étnica evidente<sup>3</sup>, representando el 15-20% de las leucemias en adultos<sup>4</sup>. Sin embargo, un estudio epidemiológico que consideró estrictamente los casos de LMC positivos para el cromosoma Ph en registros completos, reportó una incidencia de 6 a 8 casos por millón de personas por año. En Argentina, esta incidencia global se traduciría en 240 a 320 casos nuevos por año (estimación a partir de los datos de población total del censo 2010, INDEC). La edad al diagnóstico es variable, y el promedio varía según región geográfica, posiblemente debido a diferencias en la composición etaria. En Europa, el promedio de edad al diagnóstico es 60 a 65 años, mientras que en las regiones de Latinoamérica, sudeste asiático y África es menor<sup>5</sup>. La incidencia en hombres es mayor que en mujeres (cociente 1,3-1,8), y su incidencia aumenta con la edad<sup>6</sup>. No están descriptos factores etiológicos, a excepción de la exposición a altas dosis de radiación ionizante<sup>7</sup>.

#### I.1.3. Características del rearrreglo genómico *BCR-ABL1*

La translocación recíproca  $t(9;22)(q34;q11)$  origina el gen de fusión *BCR-ABL1* en el cromosoma Ph (22q-), que codifica una proteína de fusión con actividad tirosina-quinasa (otorgada por un dominio de *ABL1*) constitutivamente activa. En el 95% de los casos, los puntos de ruptura cromosómicos se pueden encontrar distribuidos a lo largo del intrón 13 o intrón 14 de *BCR* (~3kpb), y a lo largo del intrón 1 de *ABL1* (~140kpb). En un estudio en el que se caracterizaron 308 puntos de fusión *BCR-ABL1* no se encontraron puntos de fusión idénticos, sino que cada paciente albergaba una secuencia específica en la región genómica correspondiente al rearrreglo cromosómico<sup>8</sup>. Dado que las regiones intrónicas son eliminadas durante el corte y empalme del transcrito primario, los ARNm maduros son virtualmente idénticos entre todos los pacientes, difiriendo sólo en la presencia o ausencia del exón 14 de *BCR*. Estas variantes transcripcionales son denominadas b2a2 (también denominada e13a2) y b3a2 (e14a2), y codifican para una proteína de 210 kDa (p210<sup>Bcr-Abl</sup>) (Fig. 1.1).

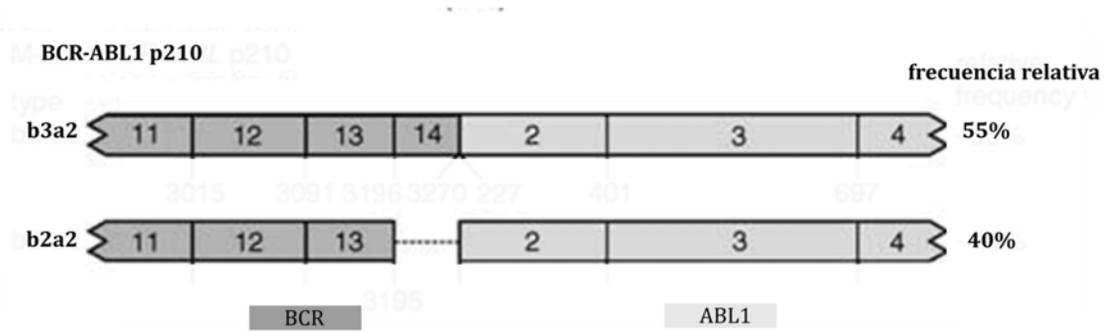


Fig. 1.1. Estructura de las isoformas de ARNm de BCR-ABL1 más frecuentes en LMC. Los números corresponden al número de exón de cada gen. Adaptado de <sup>9</sup>.

En 2/3 de los pacientes con leucemia linfoblástica aguda Ph+ y en raros casos de LMC y leucemia mieloide aguda, la ruptura en BCR ocurre en un intrón río arriba, localizado entre dos segundos exones alternativos (e2 y e2'), resultando en la isoforma e1a2, que codifica para una proteína de fusión de 190kDa (p190<sup>Bcr-Abl</sup>)<sup>2</sup>.

#### I.1.4. Diagnóstico

El diagnóstico se realiza mediante un hemograma y frotis de sangre (leucocitosis neutrofílica con precursores mieloides, eosinofilia, basofilia, plaquetas normales o aumentadas), un estudio citogenético por bandeo G en médula ósea (MO) de 20 metafases (para la detección del cromosoma Ph y de otras alteraciones citogenéticas) (Fig. 1.2), y un estudio molecular de RT-PCR en sangre periférica (SP) (para la detección del transcrito BCR-ABL1, y la caracterización de la isoforma correspondiente). En un ≈5% de los casos, el cromosoma Ph no es detectable por estudio citogenético, y la evaluación de la translocación recíproca se realiza por la técnica de FISH<sup>10</sup>. Una vez confirmado el diagnóstico se puede estratificar a los pacientes en grupos de riesgo (bajo, intermedio, alto) a través del cálculo del índice de Sokal, EUTOS o Hasford. El índice de Sokal es el más común, y utiliza como variables la edad, el tamaño del bazo, el recuento de plaquetas y el porcentaje de blastos en sangre periférica<sup>11</sup>.

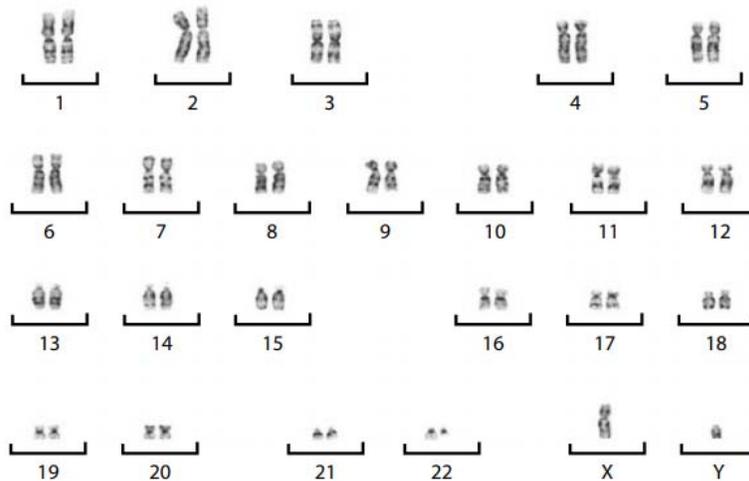


Fig. 1.2. Cariotipo de un paciente con LMC en fase crónica. La translocación recíproca  $t(9;22)(q34;q11)$  genera el cromosoma Ph (22q-) y su recíproco (9q+). Extraído de <sup>2</sup>.

### I.1.5. Fisiopatología

Las características de la proteína de fusión BCR-ABL1 están dadas por la combinación de interacciones celulares propias de BCR y ABL1, y por la aparición de propiedades nuevas, dadas por la interacción de algunos de los dominios de ambas proteínas y por pérdida de otros. El extremo N-terminal de BCR contiene un motivo *coiled-coil* esencial para la oncogenicidad de BCR-ABL1, que permite la dimerización de la proteína, la cual se demostró que es necesaria para la activación de la función quinasa de ABL1. También se conservan, de BCR, un dominio quinasa de tipo serina-treonina, dos regiones de unión a dominios (SH)<sub>2</sub> y un dominio Rho de intercambio GTP-GDP (rho-GEF). De ABL1 están presentes tres dominios de homología con Src: SH1 es el que exhibe función tirosina-quinasa, SH2 es de unión a fosfo-tirosina, y SH3 una secuencia rica en prolina de otras proteínas, como CrkL. A su vez, en ABL1, hacia el extremo C-terminal, hay un dominio de unión al ADN y otro de unión a actina <sup>12</sup>.

Las consecuencias celulares dadas por la presencia de la proteína BCR-ABL1 en células hematopoyéticas son la desregulación de varias vías de señalización, incluyendo las vías de PI3K, JAK2-STAT y MAPK. Consecuentemente, las células transformadas exhiben alta proliferación, menor adherencia al estroma de MO, menor respuesta a estímulos proapoptóticos y mayor inestabilidad genómica<sup>2</sup>. Al iniciarse en una célula madre hematopoyética, todos los linajes albergan la translocación recíproca. Sin embargo, el fenotipo se manifiesta principalmente como leucocitosis neutrofílica y trombocitosis, sin efectos evidentes en la capacidad de diferenciación de los progenitores para producir células maduras. Durante la progresión de la enfermedad, aparecen mutaciones adicionales al BCR-ABL1 que ocasionan un arresto en la diferenciación, resultando en la acumulación de blastos inmaduros.

### **I.1.6. Tratamiento**

La caracterización de la proteína BCR-ABL1 y de su rol en la patogénesis de la LMC llevó al diseño racional de drogas específicas, en particular de inhibidores competitivos del sitio de unión al ATP de la quinasa ABL1<sup>13</sup>. A principios del año 2000 se demostró que una de estas drogas, el compuesto STI571 (que luego sería patentada por Novartis como Imatinib, IM), era efectivo *in vitro* contra células leucémicas, y mostró resultados promisorios en un ensayo clínico de fase II<sup>14</sup>. El salto cuantitativo que permitió el establecimiento del IM como tratamiento de primera línea para pacientes con LMC en pacientes en fase crónica fue el ensayo de fase III llamado IRIS (*International Randomized Study of Interferon and STI571*), en el cual 1106 pacientes en 177 centros de 16 países fueron asignados de forma aleatoria al tratamiento con IM (400mg/día) o con interferón- $\alpha$  (IFN- $\alpha$ ) más citarabina, el tratamiento estándar hasta ese momento para este grupo de pacientes. A los 18 meses, la tasa de respuesta citogenética completa (0% metafases Ph+) fue del 76,2% en la rama IM vs. 14,5% en la rama IFN- $\alpha$  + citarabina. Los resultados de este ensayo a largo plazo (el último reporte tiene un seguimiento de casi 11 años) informaron que entre los pacientes asignados a la rama IM, el 48,3% completaron el tratamiento vs. sólo el 1,3% en la rama IFN- $\alpha$  más citarabina, siendo que el 65,6% de los pacientes asignados a dicho tratamiento se cambiaron a IM<sup>15</sup>. Teniendo en cuenta que para el 20,1% (111 casos) de los pacientes no había datos de supervivencia a los 10 años, la estimación de supervivencia global a los 10 años para los pacientes en la rama IM fue de 64,4 a 83,3%, considerando a los 111 pacientes muertos o vivos, respectivamente.

Sin embargo, los resultados del estudio también evidenciaron que una fracción de los pacientes no presenta respuestas adecuadas. En la rama IM se registraron 6,9% progresiones a fase acelerada o crisis blástica vs. 12,8% en la rama IFN- $\alpha$  más citarabina. En la rama IM se registraron 16,1% muertes, siendo el 56% relacionadas a la enfermedad (habiendo o no recibido trasplante de células precursoras hematopoyéticas). Respecto a la seguridad del tratamiento, el 9,3% de los pacientes en la rama IM presentaron efectos adversos serios relacionados a la medicación, en su mayoría durante el primer año de tratamiento. Se registraron 11,3% de casos de neoplasias secundarias (benignas o malignas)<sup>15</sup>.

Este ensayo clínico puede considerarse como un éxito del trabajo conjunto de la comunidad científico-médica. No sólo los resultados de respuesta de los pacientes al tratamiento fueron sustancialmente mejores a los que se habían logrado hasta el momento con terapias no dirigidas, sino que también se estableció una línea de base para la construcción de una escala internacional que permitió la estandarización del monitoreo molecular de los pacientes entre distintos laboratorios<sup>16</sup>.

### **I.1.7. Desarrollo de inhibidores de BCR-ABL1 de segunda generación**

Se estima que un 40% de los pacientes tratados con IM presentan intolerancia o resistencia a lo largo de los primeros 5 años de tratamiento, requiriendo un cambio a inhibidores de tirosina-quinasa (ITK) de segunda generación (Nilotinib o Dasatinib)<sup>17</sup>. Éstos también surgieron del diseño racional de agentes inhibidores del dominio quinasa de BCR-ABL1, y actúan, al igual que el IM, como inhibidores competitivos del ATP<sup>18,19</sup>. Ambos compuestos demostraron tanto *in vitro* como *in vivo*, mayor potencia y respuestas

moleculares más rápidas, tal que en la actualidad se encuentran aprobados para ser utilizados, al igual que IM, como primera línea de tratamiento. La decisión de cuál ITK utilizar en primera línea es determinada por el médico tratante en función de las características clínicas del paciente, objetivos del tratamiento, edad y comorbilidades, considerando el perfil de eventos adversos de cada compuesto<sup>3</sup>. Los efectos adversos están asociados a la capacidad de cada ITK de inhibir otras quinasas celulares, y con su potencia: además de BCR-ABL1, IM y Nilotinib tienen efecto sobre receptores de PDGF (*platelet-derived growth factor*) y la tirosina-quinasa KIT; el Dasatinib inhibe a múltiples quinasas, incluyendo las de la familia Src. Hasta el momento no se han detectado diferencias en la sobrevida global entre los tres compuestos<sup>3</sup>.

Son tres los mecanismos principales que han sido descritos durante la resistencia al tratamiento en células maduras: mutaciones en el dominio quinasa de *BCR-ABL1*, amplificación del gen *BCR-ABL1*, y mecanismos independientes de la actividad quinasa de *BCR-ABL1*<sup>20</sup>. En el caso de mutaciones en el dominio quinasa, a los pacientes se les indica el ITK más adecuado según la mutación que posean, siendo algunas menos sensibles a Dasatinib (V299L, T315A, F317L/V/I/C), o al Nilotinib (Y253F/H, E255K/V, F359V/I/C) (Fig.1.3). La mutación T315I ha mostrado ser resistente a los ITK de primera y segunda generación, siendo sensible sólo a Ponatinib, un ITK de tercera generación<sup>21</sup>.

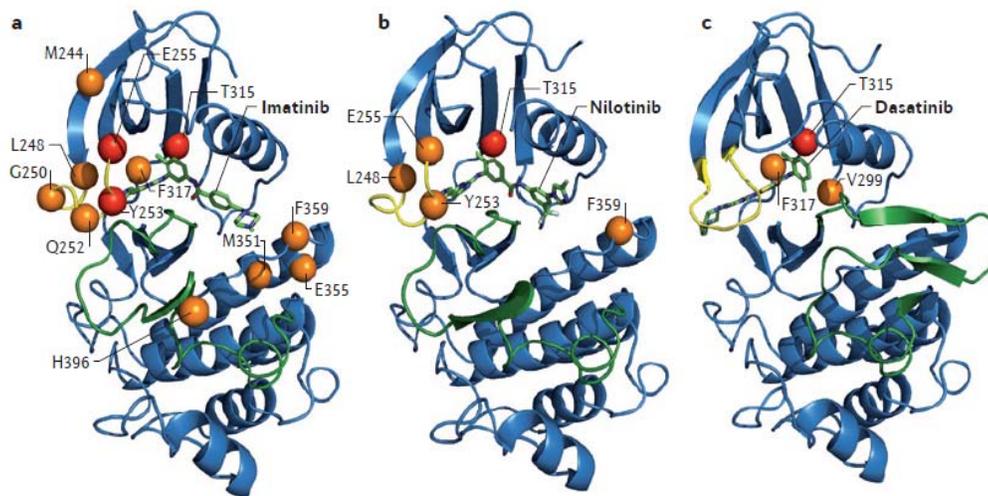


Fig. 1.3. Representación de la estructura cristalina del dominio quinasa de *ABL1*, en complejo con IM (a), Nilotinib (b) o Dasatinib (c). Se resaltan las mutaciones que confieren resistencia a cada ITK *in vitro*. En esta figura se aprecia el mayor espectro de mutaciones que pueden causar resistencia a IM en comparación con los ITK de segunda generación. La mutación T315I causa resistencia a los tres ITK. Modificado de <sup>22</sup>.

### 1.1.8. Monitoreo molecular

La respuesta al tratamiento con ITK es evaluada a nivel hematológico, citogenético y molecular. Este último consiste en la cuantificación por retrotranscripción seguida de PCR en tiempo real (RT-qPCR) de los niveles de transcritos *BCR-ABL1* en SP. El monitoreo molecular es la herramienta más sensible disponible en el contexto de la rutina clínica para la evaluación del estado de la enfermedad, principalmente de la enfermedad residual

medible. Dado que esta metodología tiene una alta variabilidad<sup>23</sup>, los valores de transcritos *BCR-ABL1* deben ser informados como cociente (número de copias *BCR-ABL1*/número de copias del gen control) en la escala internacional (IS)<sup>24</sup>. Dicha escala fue establecida para facilitar la comparabilidad de los resultados entre distintos laboratorios, asignándole a la mediana de los valores de %*BCR-ABL1*/gen control detectados al diagnóstico en 30 de los pacientes que participaron del ensayo clínico IRIS un valor de 100%, tal que las cuantificaciones posteriores sean informadas como reducciones (en escala logarítmica) respecto a dicha línea de base<sup>16</sup>. Es decir, que un paciente que presente niveles de transcritos *BCR-ABL1* 1000 veces menores a los observados al diagnóstico en los pacientes del ensayo IRIS, se le asignaría una respuesta molecular (RM) de 3 (3 log de reducción), o %*BCR-ABL1*<sup>IS</sup>=0,1% (Fig. 1.4).

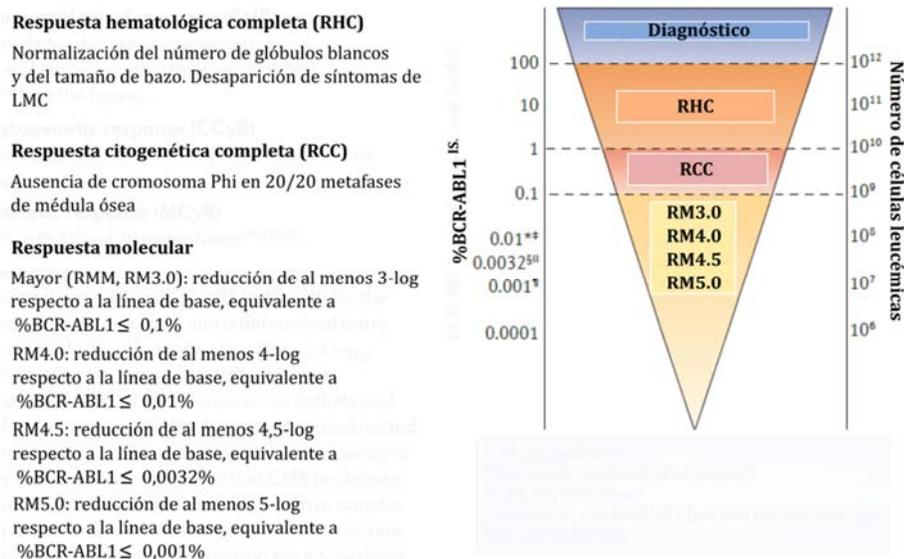


Fig. 1.4. Equivalencia entre los niveles de respuesta hematológica, citogenética y molecular. Modificado de <sup>22</sup>.

La importancia de la cuantificación de los transcritos *BCR-ABL1* en pacientes tratados con ITK radica en que los niveles de RM alcanzada a determinados tiempos post-tratamiento son indicadores pronósticos de la evolución del paciente. Los criterios de la *European Leukemia Net* para la clasificación de la respuesta al tratamiento con ITK en 1<sup>ra</sup> línea tienen en cuenta el %*BCR-ABL1*<sup>IS</sup> a los 3, 6 y 12 meses, en conjunto con el porcentaje de metafases Ph<sup>+</sup> en MO <sup>24</sup> (Tabla 1.1).

Tabla 1.1: definiciones de respuesta a los ITK en primera línea (fase crónica, fase acelerada y crisis blástica) de la European Leukemia Net 2013.

Tiempo	Respuesta óptima	Advertencia	Falla
<b>Basal</b>		Alto riesgo Ruta mayor ACA/Ph <sup>+</sup>	
<b>3 meses</b>	BCR-ABL <sup>IS</sup> ≤ 10% y/o Ph <sup>+</sup> ≤ 35% (RCP)	BCR-ABL <sup>IS</sup> > 10% y/o Ph <sup>+</sup> 36-95%	No RHC y/o Ph <sup>+</sup> > 95%
<b>6 meses</b>	BCR-ABL <sup>IS</sup> ≤ 1% y/o Ph <sup>+</sup> 0% (RCC)	BCR-ABL <sup>IS</sup> 1-10% y/o Ph <sup>+</sup> 1-35%	BCR-ABL <sup>IS</sup> > 10% y/o Ph <sup>+</sup> > 35%
<b>12 meses</b>	BCR- ABL <sup>IS</sup> ≤ 0,1% (RMM)	BCR-ABL <sup>IS</sup> 0,1-1%	BCR-ABL <sup>IS</sup> > 1% y/o Ph <sup>+</sup> > 0%
<b>A partir de entonces, en cualquier momento</b>	RMM o mejor	ACA/Ph <sup>-</sup> (-7, o 7q-)	Pérdida de RHC Pérdida de RCC Pérdida de RMM**, confirmada Mutaciones ACA/Ph <sup>+</sup>

\*\*en dos pruebas consecutivas, una de ellas ≥ 1% IS; ACA/Ph<sup>+</sup>: alteraciones cromosómicas clonales en células Ph<sup>+</sup>; ACA/Ph<sup>-</sup>: alteraciones cromosómicas clonales en células Ph<sup>-</sup>; RMM: respuesta molecular mayor; RHC: respuesta hematológica completa; RCC: respuesta citogenética completa; RCP: respuesta citogenética parcial

Inicialmente, la única forma de armonizar la metodología de RT-qPCR a la escala internacional consistía en el intercambio de muestras de pacientes con un laboratorio central de referencia (que había participado del establecimiento de la línea de base del ensayo IRIS) para la obtención de un factor de conversión a la escala IS; lo cual resultó en una mejoría en las tasas de concordancia inter-laboratorio<sup>25</sup>. Sin embargo, se trata de un proceso laborioso y costoso, por lo tanto son pocos los laboratorios que han tenido acceso a dicha estandarización. El desarrollo reciente del primer panel de referencia genética para *BCR-ABL1* (calibradores primarios), validado por la Organización Mundial de la Salud ha surgido como una respuesta a dicha limitación. Se trata de material de referencia consistente en 4 calibradores; cada calibrador tiene asignado un valor de cociente *BCR-ABL1*/gen control en la IS, el cual fue obtenido por múltiples mediciones en laboratorios estandarizados de referencia<sup>26</sup>. Dado que el stock de estándares primarios es limitado, su accesibilidad está restringida a fabricantes de kits o de estándares secundarios. En nuestra región, la implementación de un programa de armonización a la IS con calibradores secundarios de desarrollo local ha permitido un aumento sustancial en la cantidad de laboratorios estandarizados. Estos calibradores fueron desarrollados en nuestro laboratorio, y consisten en mezclas de las líneas celulares K-562 (BCR-ABL<sup>+</sup>) y HL-60 (BCR-ABL<sup>-</sup>) en distintas proporciones, las cuales representan 4 niveles de %BCR-ABL<sup>IS</sup>; las mezclas fueron liofilizadas y distribuidas a laboratorios de Argentina y otros países de Latinoamérica<sup>27</sup>. Al día de hoy, el programa incluye 29 laboratorios, de los cuales 19 se encuentran en Argentina, y los restantes están distribuidos entre Uruguay, Paraguay, Colombia, Panamá, Perú, Costa Rica, Guatemala y México. Este programa no sólo permitió la accesibilidad del material, sino también la conformación de una red de laboratorios estandarizados para el intercambio de protocolos, la evaluación de parámetros relacionados a la calidad de las mediciones, y la colaboración en proyectos de investigación.

### **I.1.9. Discontinuación de ITK**

En pacientes respondedores a los ITK, los niveles de transcritos *BCR-ABL1* disminuyen a lo largo del tiempo, obteniendo RM profundas (definidas como RM4.0, RM4.5 y RM5.0). En los casos donde no se detectan transcritos *BCR-ABL1*, algunos autores utilizan el término "enfermedad mínima residual indetectable" (UMRD). La estimación a 5 años de tratamiento con IM es que hasta un 50% de los pacientes puede alcanzar una UMRD <sup>28</sup>. Ésto ha llevado a la ejecución de numerosos ensayos clínicos de discontinuación del tratamiento, en los que de forma consistente se observó que alrededor de un 40-60% de los pacientes pueden mantenerse en remisión libre de tratamiento (TFR) <sup>29,30</sup>. En estos ensayos, los pacientes son seleccionados bajo estrictos criterios, que, aunque pueden variar ligeramente, en general incluyen:

- haber logrado una respuesta de al menos RM4.5, y mantenerla durante al menos 24 meses (pudiendo variar entre RM4.0-RM5.0);
- una duración total del tratamiento con ITK de al menos 5 años
- 18 años de edad
- no haber presentado resistencia a ITK de 1<sup>ra</sup> línea
- transcritos típicos (b2a2 o b3a2) al diagnóstico
- monitorearse de forma regular al discontinuar el tratamiento, en un laboratorio estandarizado (1 vez el mes durante los primeros 6 meses, cada 6 semanas los siguientes 6 meses, cada 3 meses a partir del segundo año).

Una de las características más llamativas de estos ensayos es que en la mayoría de los pacientes en los que ocurre una recaída molecular (definida como pérdida de RM3.0/MMR en SP), sucede en los primeros 6-7 meses post-discontinuación, y el 100% vuelve a responder al tratamiento con ITK, sin presentar mutaciones. En una revisión del 2016 se reportó sólo 1 caso de progresión a crisis blástica 8,5 meses después de reiniciar el tratamiento con IM <sup>30</sup>. Estos resultados son muy promisorios; sin embargo, si se tiene en cuenta el porcentaje de pacientes que alcanzan los requisitos para discontinuar el tratamiento, sólo un 10% de los pacientes tratados con IM podrá mantenerse en TFR <sup>17</sup>, con lo cual la mayoría de los pacientes deben mantener el tratamiento con ITK de forma indefinida. Ésto tiene implicancias en la calidad de vida de los pacientes, dados los efectos adversos de los ITK, y en el sistema de salud, dado que el precio de los ITK es sustancialmente alto.

### **I.2. Células madre hematopoyéticas**

Las células madre hematopoyéticas (HSC, *hematopoietic stem cells*) se definen por su capacidad de autorrenovación y de diferenciación multipotente, lo cual es posible a través de divisiones asimétricas, generando una célula hija diferenciada y otra HSC. La célula hija diferenciada sale del reservorio de HSC y genera células maduras a través de una cascada de pasos de diferenciación. Las HSC también tienen ciclos de autorrenovación por divisiones simétricas, en los cuales generan HSC hijas capaces a su vez de autorrenovarse y diferenciarse. La elección entre divisiones simétricas y asimétricas está regulada por

múltiples factores, tanto intrínsecos (por ejemplo, alteraciones genéticas) como extrínsecos (por ejemplo, señales del microambiente de la MO)<sup>31-33</sup>. En su conjunto, estos mecanismos permiten la expansión de una única HSC en un clon de muchas células, y el mantenimiento del sistema hematopoyético a lo largo de la vida del organismo<sup>34</sup>.

La fracción de células Lin-CD34<sup>+</sup>CD38<sup>-/low</sup> contiene a las HSC y progenitores multipotentes (MPP, *multipotent progenitors*), los cuales tienen el potencial de generar progenitores comprometidos a linaje mieloide (CMP, *common myeloid progenitors*) o linfóide (MLP, *multi-lymphoid progenitors*). A su vez, los CMP pueden generar progenitores de tipo granulocito-macrófago (GMP) o megacariocito/eritrocito (MEP). Los progenitores comprometidos a linaje están contenidos en la fracción de células Lin-CD34<sup>+</sup>CD38<sup>+</sup> (Fig. 1.5)<sup>35</sup>. Las células terminalmente diferenciadas (linfocitos, células *Natural Killer*, células dendríticas, granulocitos, megacariocitos, eritrocitos) son células con un ciclo de vida corto, y por lo tanto el organismo requiere la regeneración constante de nuevas células a partir de los precursores presentes en la MO. En modelos murinos se ha demostrado la existencia de un *pool* de células quiescentes, las cuales se considera que no contribuyen al mantenimiento diario de la hematopoyesis, sino que funcionarían como reservorio frente a situaciones de alta demanda, por ejemplo, frente a una lesión o proceso inflamatorio<sup>36</sup>. De forma análoga, en humanos, la evaluación del ciclo celular por citometría de flujo de células provenientes de MO de personas sanas indica que la mayoría de las HSC se encuentran en la fase G<sub>0</sub><sup>35</sup>, es decir, quiescentes, mientras que estimaciones indirectas del ciclo de vida indican que las HSC humanas se dividen cada 175-350 días<sup>37</sup>. A este modelo clásico se ha agregado una capa de complejidad otorgada por la heterogeneidad del reservorio de HSC: no todas las HSC tienen la misma capacidad de autorrenovarse ni pueden generar clones del mismo tamaño. Ésto implica que el ciclo de vida puede variar entre las distintas subpoblaciones de HSC<sup>34</sup>.

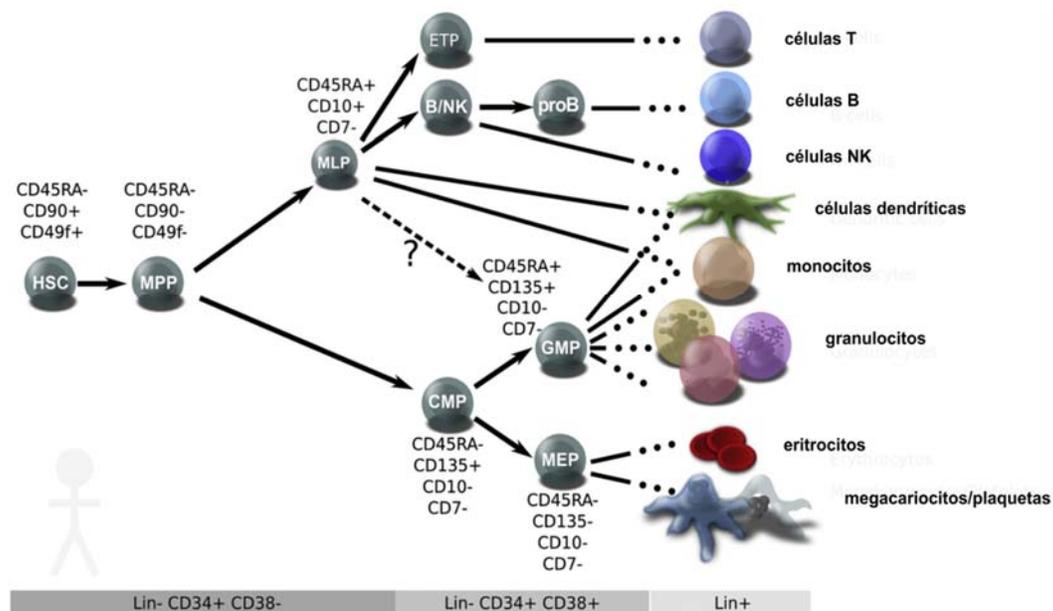


Fig.1.5. Esquema del sistema hematopoyético humano. Se indican los marcadores fenotípicos utilizados para diferenciar cada subpoblación. Modificado de <sup>35</sup>.

### I.2.1. Modelos para estudiar la funcionalidad de las HSC

El ensayo sustituto (*surrogate assay*) de referencia para las HSC es el modelo de trasplantes seriales de HSC en ratones que fueron sometidos a dosis subletales de radiación para condicionar la MO. Para ello, se inyectan cantidades limitantes de HSC en hospedadores primarios. Si estas células se autorrenovaron en el hospedador, entonces tendrán la capacidad de implantarse al ser inyectadas en hospedadores secundarios, y así sucesivamente en hospedadores terciarios, etc. Para el caso de HSC humanas, se utilizan ratones inmunodeficientes (NOD-SCID: *nonobese diabetic-severe combined immunodeficiency*, o NSG: *NOD-SCID IL2 $\gamma$ <sup>null</sup>*) que permiten la implantación de xenotrasplantes humanos. Mediante este tipo de ensayos, las poblaciones Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>low</sup> y Lin<sup>-</sup>CD34<sup>+</sup>CD38<sup>-</sup> provenientes de sangre de cordón humana fueron caracterizadas como HSC, capaces de generar células del linaje mielóide, linfóide y eritroide, aunque con heterogeneidad en su capacidad de autorrenovación; a su vez, se observó que la población CD38<sup>-</sup> está enriquecida unas 40 veces en HSC respecto a la población CD38<sup>low</sup> <sup>34</sup>.

Alternativas al ensayo *in vivo* son los ensayos de largo término *in vitro*, LTC-IC (*Long-term culture-Initiating cell*), en los cuales las células provenientes de MO o SP pueden utilizarse como células iniciadoras de un cultivo en presencia de células proveedoras de factores de crecimiento (*feeder layers*), las cuales intentan simular las condiciones presentes en el microambiente de la MO. Estos cultivos tienen una extensión mínima de 5 semanas, y permiten evaluar, de forma retrospectiva, la presencia de células primitivas en el inóculo inicial.

### I.2.2. Origen y persistencia del clon leucémico

Los primeros indicios de que la LMC se origina en una HSC transformada provienen de casos en los que, luego de realizar una transfusión de sangre proveniente de pacientes con LMC a receptores con neutropenia severa, se detectó implantación en MO y progenie Ph<sup>+</sup><sup>38</sup>. Experimentos realizados en modelos murinos transgénicos, en los que se inducía la expresión de *BCR-ABL1* en células progenitoras, demostraron que *BCR-ABL1* no era capaz de otorgar capacidad de autorrenovación a dichas células, descartando un origen del clon leucémico en esta fracción<sup>39</sup>. En cambio, si la expresión de *BCR-ABL1* era inducida en la fracción de células más primitivas, los ratones desarrollaban una enfermedad tipo LMC<sup>40</sup>. Por otro lado, en pacientes con LMC, se pueden encontrar células Ph<sup>+</sup> en todos los linajes<sup>41</sup>, sugiriendo que la translocación ocurrió en una célula multipotente. Se denomina entonces a la población responsable del origen y mantenimiento del clon leucémico como células madre leucémicas (LSC), consideradas células multipotentes con capacidad de autorrenovación.

Estudios recientes han cuestionado que *BCR-ABL1* sea suficiente para inducir la enfermedad, ya que la expresión de una sola copia de *BCR-ABL1* mediante experimentos de *knock-in*, río abajo del promotor endógeno de *BCR*, no causó el desarrollo de leucemia en los ratones, aunque sí se detectó un aumento en la capacidad de implantación en MO<sup>42</sup>. Otros indicios de que *BCR-ABL1* es necesario pero no suficiente para el desarrollo de la enfermedad son los registros de sobrevivientes de la bomba atómica en Japón, en los que se observó un período de latencia de alrededor de 7 años<sup>43</sup>. Por otro lado, en personas

sanas y líneas celulares no relacionadas a LMC también pueden detectarse, en bajos niveles, transcritos *BCR-ABL1*, sugiriendo que son necesarias alteraciones genéticas (o epigenéticas) adicionales para conferir a dichas células capacidad leucemogénica<sup>44</sup>.

### I.2.3. Características de las LSC

El caso de la LMC es, junto a la leucemia mieloide aguda y la leucemia linfoblástica aguda, uno de los ejemplos más claros del modelo de célula *stem* tumoral (CSC, *cancer stem cell*). Las CSC fueron inicialmente definidas como un subgrupo de células biológicamente distintas del resto del tumor, con la capacidad exclusiva de perpetuar el crecimiento del tumor de forma indefinida<sup>45</sup>. Definiciones más recientes consideran que no es necesario hacer inferencias respecto a la célula de origen donde ocurrió la transformación, y excluye a las células pre-malignas<sup>46</sup>. Esta definición permite englobar a poblaciones de CSC que dan evidencias de ser heterogéneas, conteniendo subclones con distinta capacidad de autorrenovación e interacción con señales del microambiente.

Una de las limitaciones para el estudio de las CSC o LSC es que los marcadores fenotípicos disponibles para su aislamiento no reemplazan a los ensayos funcionales, ya que los marcadores pueden verse alterados por el mismo proceso tumoral o leucémico. Los ensayos funcionales también presentan sus limitaciones ya que, tanto *in vivo* como *in vitro*, las condiciones en general han sido optimizadas para evaluar la capacidad *stem* de poblaciones normales, y tienen una duración temporal menor a los ciclos de vida de las células *stem* adultas humanas.

El hecho de que el 60% de los pacientes con LMC en los que se discontinuó el tratamiento con ITK presenta recaída molecular en los primeros meses post-discontinuación, sugiere que el clon leucémico puede persistir y mantener su capacidad proliferativa, a pesar de haber estado bajo tratamiento con ITK por muchos años. El paradigma actual considera que las responsables de la persistencia son las LSC, por varias evidencias:

- pueden ser detectadas en MO de pacientes con RM profunda<sup>47,48</sup>;
- en los pacientes que presentan recaída molecular luego de la discontinuación, la cinética de proliferación del clon leucémico es similar a la de un típico clon leucémico al diagnóstico<sup>49</sup>;
- son capaces de sobrevivir al tratamiento con ITK *in vitro*, por un mecanismo independiente de la actividad quinasa de BCR-ABL1<sup>50,51</sup>;
- exhiben menor sensibilidad al tratamiento con ITK *in vitro* que la fracción de células progenitoras CD34<sup>+</sup>CD38<sup>+</sup> <sup>52</sup>.

Entre las evidencias provenientes de ensayos *in vitro* que intentan describir los mecanismos por los cuales estas células son resistentes a los ITK se encuentran: la quiescencia de una sub-población de células CD34<sup>+</sup> luego del tratamiento con ITK, aún en presencia de factores de crecimiento<sup>53</sup>; la protección de la apoptosis inducida por ITK otorgada por el co-cultivo con células *stem* mesenquimales, a través de un mecanismo mediado por receptores de proteínas Wnt<sup>54</sup>; y la expresión de mayores o menores niveles de *BCR-ABL1* respecto a las células más diferenciadas<sup>52,55</sup>.

Estas observaciones han llevado a la búsqueda de blancos específicos de las LSC, vías independientes de la actividad quinasa de BCR-ABL1, y estrategias terapéuticas que puedan complementar el tratamiento con ITK, como son inhibidores de las vías de JAK2 y Hedgehog; inhibidores del proceso de autofagia; reguladores epigenéticos; inhibidores de ALOX5, y anticuerpos monoclonales para IL1-RAP. También se han intentado estrategias para interferir la interacción con el nicho de la médula ósea, por ejemplo, inhibidores de CXCR4 e inhibidores del checkpoint inmune PD-1/PD-L1<sup>17</sup>.

### **I.3. Regulación de la expresión génica mediada por microARNs**

Los microARNs son ARNs pequeños (19-25 nucleótidos) no codificantes, que actúan como reguladores de la expresión génica. El primer microARN descrito fue lin-4 en el nematodo *C. elegans*, en el cual se demostró que es esencial durante el proceso de desarrollo, mediante un mecanismo dependiente de la unión al extremo 3'UTR del transcripto lin-14<sup>56,57</sup>. Trabajos posteriores demostraron la existencia de microARNs en distintos grupos de organismos, como así también su alto grado de conservación evolutiva. En humanos, se han descrito unos 519 microARNs para los cuales hay una alta confianza de que sean reales, y se estima que un 60% de las regiones 3'UTR de ARNm humanos contienen sitios conservados de unión a microARNs<sup>58</sup>. Si se consideran todos los registros de microARNs en la versión más reciente de miRBase (v21), llegan a 1881 microARNs.

Entre otros ARNs no codificantes, los microARNs han resultado de especial interés por poseer características de relevancia funcional<sup>59</sup>, como son:

- poseen un alto grado de conservación evolutiva (hay evidencias de microARNs hasta en anémonas de mar y esponjas);
- la diversidad de microARNs correlaciona con la complejidad genómica;
- son moléculas relativamente estables, pudiendo ser detectados en fluidos como sangre y orina;
- sus efectos son pleiotrópicos, ya que un mismo microARN puede regular varios ARNm blanco;
- muestran patrones de expresión específicos de tejido y subtipo celular.

#### **I.3.1. Organización genómica**

Muchos microARNs forman parte de *clusters* genómicos a partir de los cuales se generan transcriptos policistrónicos. Los *loci*, formando parte de *clusters* o como unidades independientes, pueden estar localizados en regiones intergénicas (en ese caso poseen su propio promotor), o estar incluidos en intrones o exones de otros genes. Una revisión reciente de todos los microARNs registrados en miRBase (v21) clasificó al 76% de las secuencias de microARNs como intragénicas, de las cuales el 64% se encontraba en intrones de otros genes (codificantes o no codificantes)<sup>60</sup>. A su vez, los microARNs son agrupados en familias en base a las características de su región semilla, tal que aquellos microARNs con similitud de secuencia tendrán un espectro de ARNm blancos similares en secuencia. Los miembros de una familia pueden o no estar localizados en un mismo *cluster* genómico.

### I.3.2. Biogénesis de microARNs

Los microARNs pueden generarse a partir de mecanismos canónicos o no canónicos. Durante la biogénesis de microARNs canónicos, los transcriptos primarios (pri-microARNs) son sintetizados por la ARN polimerasa II, con lo cual poseen 5'CAP y pueden o no estar poli-adenilados en su extremo 3'. Los pri-microARNs forman estructuras de tipo *hairpin*, que son clivadas por el complejo Microprocesador (Drosha-DGCR8) para formar los pre-microARNs, de 70nt, también con estructura de *hairpin*. Estos precursores son exportados al citoplasma mediante un complejo que involucra a Exportina-5, en un proceso dependiente de Ran-GTP<sup>61</sup>. En el citoplasma son procesados por un complejo enzimático que contiene a la endonucleasa Dicer, liberando un ARN doble cadena de 18-25pb, típicamente con 2 nucleótidos sin aparear en ambos extremos 3', y un grupo 5' fosfato en cada extremo. El dúplex es cargado en el complejo RISC (*RNA-induced silencing complex*) a través de una proteína de la familia Argonauta (Ago), generalmente a través de Ago2, y una de las hebras podrá actuar como hebra guía. La otra hebra, denominada microARN\*, es liberada del complejo y degradada (Fig 1.6, A). Dependiendo si el microARN maduro se genera a partir del brazo 5' o 3' del pre-microARN, se le agrega al nombre el sufijo 5p o 3p, respectivamente. Entre los microARNs de biogénesis canónica también se incluye a aquellos generados a partir de transcriptos primarios de otros genes, ya sea codificantes o no codificantes, pero que, al igual que los que provienen de una unidad génica independiente, dependen del complejo Microprocesador para la generación de los pre-microARN con estructura de *hairpin* (Fig. 1.6, B).

Los microARNs de biogénesis no canónica son, por definición, aquellos que no requieren el procesamiento por el complejo Microprocesador o por Dicer. Por ejemplo, algunos microARNs generados a partir del corte y empalme de intrones (mirtrones) ingresan a la vía como pre-microARNs (Fig. 1.6, C). Otros consisten en *short-hairpin* ARNs (shARNs), siendo los más caracterizados aquellos producidos por la ARNpol II, por ejemplo, algunos derivados de las regiones 5' de ARNm o de los transcriptos antisentido de éstos. Finalmente, una tercera clase de microARNs de biogénesis no canónica se transcriben en tándem con o como parte de otro tipo de ARN pequeño, por ejemplo ARNt, ARNs pequeños nucleares, ARNs pequeños nucleolares. La mayoría de los microARNs de caracterización reciente se clasifican como microARNs de biogénesis no canónica, en su mayoría, mirtrones.

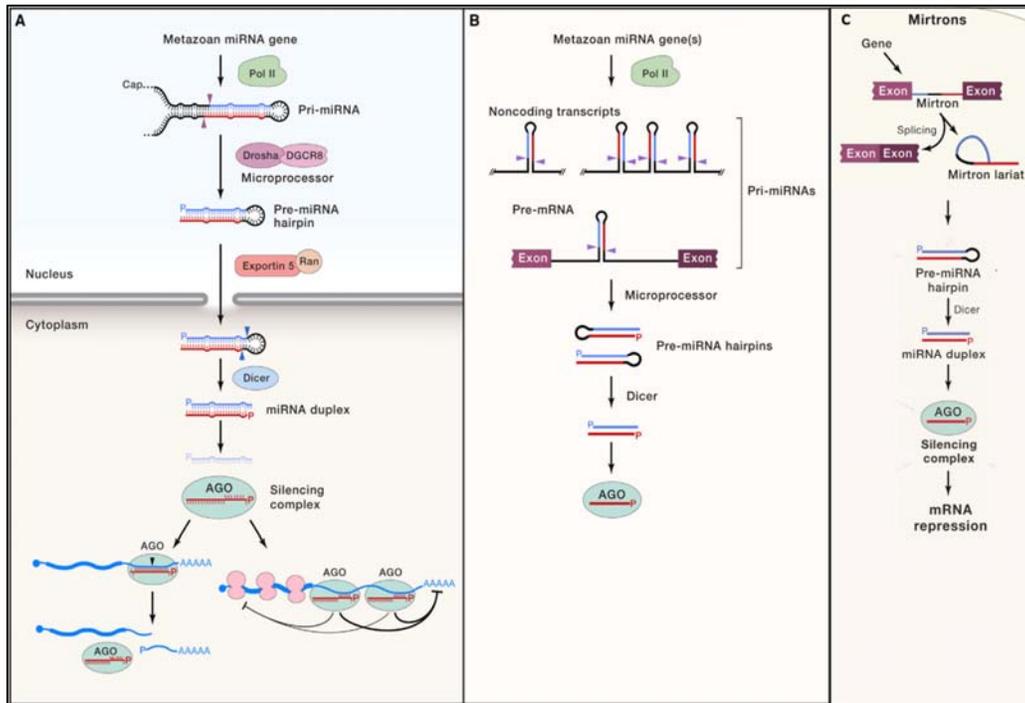


Fig.1.6. Biogénesis de microARNs canónicos (A-B) y no canónicos (C). Para los no canónicos se muestra el caso mejor estudiado, correspondiente a la generación de microARNs a partir del corte y empalme de intrones de otros genes (mirtrones). Modificado de <sup>58</sup>.

### 1.3.3. Mecanismos de acción de microARNs

A diferencia del mecanismo de silenciamiento utilizado en experimentos de *knock-down* (siARN) o a lo que ocurre con los microARNs en plantas, la mayoría de los microARNs animales disminuyen los niveles del ARNm blanco mediante un mecanismo de decaimiento (*decay*) o represión de la traducción. La unión de los microARNs a los ARNm ocurre, generalmente, a través de su región semilla, correspondiente a los nucleótidos 2-7 a partir del extremo 5', con regiones de distinto grado de complementariedad en la zona regulatoria 3'UTR de los ARNm. Una vez formado el complejo, se reclutan proteínas adaptadoras que median la interacción con la proteína PABPC (*poly(A)-binding protein*), asociada a la cola de poli-A de los ARNm, y proteínas deadenilasas que provocan el acortamiento de la cola de poli-A, con la consecuente desestabilización del ARNm. El reclutamiento de proteínas adaptadoras también disminuye la eficiencia de traducción del ARNm (Fig. 1.6, A)<sup>58</sup>.

También se han reportado mecanismos de acción no canónicos de los microARNs, como silenciamiento a través del reclutamiento de proteínas remodeladoras de la cromatina<sup>62</sup>, e inducción de la traducción a través de la unión al 5'UTR de ARNm de proteínas ribosomales<sup>63</sup>.

### 1.3.4. Participación de microARNs en procesos biológicos

La generación de ratones *knock-out* para microARNs ha demostrado la relevancia funcional de estas moléculas, ya que experimentos de pérdida de función para 20 de las 27 familias de microARNs más conservadas en animales, resultaron en fenotipos alterados.

Entre los fenotipos observados se observaron: viabilidad embrionaria reducida, enfermedades neurológicas graves, infertilidad, ceguera, sordera, desórdenes del sistema inmune y cáncer. Es importante considerar que, dada la redundancia entre microARNs pertenecientes a la misma familia, varios de estos fenotipos sólo se expresaron al anular varios de los miembros de la familia en simultáneo<sup>58</sup>.

En particular, respecto al sistema hematopoyético, existen múltiples evidencias de su participación en los procesos de hematopoyesis normal y maligna, incluyendo la comisión a linaje, diferenciación, apoptosis, y funcionalidad de células inmunes efectoras<sup>64</sup>. En ratones transgénicos *knock-out* condicionales para Dicer, en los que dicha proteína estaba ausente sólo en células del linaje hematopoyético, se observó mediante ensayos funcionales *in vivo* una reducción en la población de células primitivas y células maduras<sup>65</sup>. Asimismo existen múltiples trabajos que describen la expresión desregulada de microARNs en neoplasias hematológicas, y entre los mecanismos intervinientes se han descrito anomalías cromosómicas (por ej. en Leucemia linfocítica crónica, linfoma de células B, leucemia linfoblástica aguda), cambios en el control transcripcional (por ej. a través de c-Myc, p53, PU.1, C/EBP $\alpha$ ), modificaciones epigenéticas (por ej. miR-223 en leucemia mieloide aguda) y defectos en la maquinaria de biogénesis de los microARNs (por ej. por mutaciones inactivantes o cambios en la expresión de los componentes)<sup>66</sup>.

#### **I.4. Antecedentes de la hipótesis**

La mayor parte de las evidencias respecto al efecto de los ITK en las LSC fueron obtenidas en experimentos realizados *in vitro*, con muestras provenientes de pacientes con LMC al diagnóstico. Las evidencias obtenidas de pacientes tratados con ITK por períodos prolongados, por el contrario, son limitadas. Entre ellas se encuentran algunos reportes de persistencia de LSC en pacientes bajo tratamiento con Imatinib, Nilotinib y Dasatinib, aún cuando éstos alcanzaron una RM profunda sostenida en el tiempo <sup>48,67</sup>. Por otro lado, dado que la respuesta de los pacientes al tratamiento con IM es heterogénea, siendo que un 40% debe rotar a inhibidores de segunda línea, se desconoce si dicha respuesta diferencial tiene alguna relación con las propiedades de las LSC al diagnóstico, o que pudieran emerger durante el tratamiento.

Adicionalmente, en el contexto de la discontinuación del tratamiento con ITK, se desconoce si los niveles de LSC pueden servir como factor pronóstico de la capacidad de lograr el TFR, o, en aquellos que sí lo lograron, si la presencia de LSC funcionales representa un riesgo de recaída futura.

Por otro lado, las LSC han sido extensamente estudiadas en LMC y otras neoplasias hematológicas. Sin embargo, los ensayos clínicos con compuestos dirigidos a esta población no han logrado buenos resultados hasta el momento, siendo que la mayoría de ellos ha presentado una alta toxicidad. En este sentido, aún no se comprenden los mecanismos que permiten a una LSC tener características asociadas a un proceso oncogénico, como es mayor proliferación, desregulación de la interacción con la matriz extracelular, menor sensibilidad a señales apoptóticas, y aún así seguir comportándose como células primitivas, con capacidad de autorrenovación y multipotencialidad.

## **HIPÓTESIS**

La hipótesis general del trabajo es que la carga de células madre leucémicas (LSC) en pacientes con leucemia mieloide crónica (LMC) es indicativa de la respuesta al tratamiento con inhibidores de tirosina-quinasa (ITK) y de la posibilidad de discontinuar la terapia de forma segura. Sus características estarían dadas, al menos en parte, por mecanismos de regulación génica mediados por un conjunto de microARNs diferencialmente expresados respecto de sus contrapartes normales, las células madre hematopoyéticas (HSC).

## **OBJETIVOS GENERALES**

- Evaluar el impacto clínico de la carga de LSC en pacientes con LMC al diagnóstico y bajo tratamiento con ITK
- Identificar mecanismos moleculares diferenciales a través de la evaluación de los niveles de microARNs en las LSC en comparación con las HSC de pacientes y dadores sanos.

## **OBJETIVOS ESPECÍFICOS**

- 1) Cuantificar la carga leucémica en la fracción primitiva y progenitora en muestras de pacientes con LMC al diagnóstico, y bajo tratamiento con ITK, utilizando ensayos funcionales *in vitro*.
- 2) Correlacionar la carga de LSC con la cinética y profundidad de respuesta a los ITK.
- 3) Caracterizar el miRnoma de LSC y HSC de pacientes con LMC al diagnóstico, en comparación con dadores sanos.

## CAPÍTULO II. MATERIALES Y MÉTODOS

### II.1. Muestras de pacientes.

El proyecto fue aprobado por el Comité de ética del Instituto Alexander Fleming. Los pacientes firmaron un consentimiento informado elaborado especialmente para este proyecto. Se obtuvieron muestras de SP o MO de pacientes con LMC en fase crónica, de dos tipos: al diagnóstico (previo tratamiento), o a distintos puntos de tiempo de tratamiento con ITK (Imatinib o Nilotinib). Los pacientes fueron clasificados según su respuesta al tratamiento de acuerdo a los indicadores sugeridos por la *European Leukemia net*<sup>24</sup> (ver Introducción, tabla 1.1). En el caso de muestras al diagnóstico, 2-3mL de SP fueron utilizados para la evaluación de transcriptos *BCR-ABL1* por RT-PCR cualitativa para la isoforma p210 (protocolo detallado en la sección II.5).

### II.2 Aislamiento de la fracción CD34+.

Se separó la fase enriquecida en células mononucleares (MNC) mediante centrifugación en gradiente de densidad (Ficoll-Paque PLUS, GE Healthcare Life Sciences) durante 30min a 480xg. Los MNC se obtuvieron a partir de la recolección de la interfase. Se realizó un lavado en PBS (GIBCO), lisis de glóbulos rojos a 37°C (EDTA 0,13mM, KHCO<sub>3</sub> 1 mM, NH<sub>4</sub>Cl 170mM, pH 7,3), y remoción del plasma enriquecido en plaquetas mediante centrifugación a baja velocidad (12-15 minutos a 200xg). Se utilizaron hasta 2x10<sup>8</sup> de MNC para el aislamiento de la fracción CD34+, el cual consistió en una selección positiva utilizando *beads* magnéticas (CD34 MicroBeads, Miltenyi Biotech). El protocolo consistió en 30min de incubación de 100µl de *beads* y 100µl de bloqueante de receptores FcR por cada 1x10<sup>8</sup> MNC (en ~300µl de buffer MACS frío: BSA 0,5% en PBS/EDTA 2mM, pH 7,2), en un baño a 2-8°C. Se realizaron 2 lavados con 4mL de buffer MACS. Para el sembrado en la columna, lavado y elución, se utilizó buffer MACS frío. Las células fueron sembradas en una columna MS (Miltenyi Biotech) en 500µL, lavadas 4 veces (1x 1mL, 3 x 500µL), y eluidas en 2mL. La fracción CD34+ fue criopreservada en 1mL de medio de congelación (DMEM 50%/albúmina sérica humana 40%/DMSO 10%). En los casos en los que se disponía de suficiente número de células, se utilizaron 10.000 células para evaluar la pureza de la fracción mediante citometría de flujo (ver sección II.6.1 para más detalles).

### II.3. Ensayo de corto término para células progenitoras (CFU).

Las células CD34+ criopreservadas fueron descongeladas en 20mL de DMEM (GIBCO). Luego de contar células viables con el colorante vital Azul Tripán (GIBCO) en cámara de Neubauer, se plaquearon 500 células por placa, resuspendidas en 100µl de IMDM (GIBCO) o RPMI (RPMI 1640, GIBCO), en placas de cultivo p35 conteniendo 1mL de metilcelulosa enriquecida en factores de crecimiento (Methocult H4435 Medium, Stem Cell Technologies), por duplicado. Para ello, las células fueron agregadas a 1,1mL de medio semisólido, se vortexearon, y luego de 5min a TA se plaquearon en p35 utilizando una jeringa de 1mL sin aguja. Los factores incluidos en este medio son: SCF (*Stem cell factor*), IL-3 (*Interleukin-3*), IL-6 (*Interleukin-6*), EPO (*erithropoietin*), G-CSF (*granulocyte colony-stimulating factor*) y GM-CSF (*granulocyte-macrophage colony-stimulating factor*), todos como proteínas recombinantes humanas. El medio también incluye SFB, IMDM, 2-Mercaptoetanol y suplementos. Las células fueron incubadas en estufa a 37°C, en atmósfera húmeda con CO<sub>2</sub> 5%, en el interior de una cámara húmeda (placa p100 estéril

conteniendo 1 p35 con agua estéril por cada 2 p35 con células) durante 14-18 días. Al cabo de ese tiempo, las colonias fueron contabilizadas al microscopio óptico, clasificadas según el tipo de progenitor hematopoyético (progenitores macrófago-granulocito: CFU-GM/CFU-M/CFU-G, eritroides: BFU-E/CFU-E, de linaje múltiple: CFU-GEMM) (Fig. 2.1), y extraídas de la metilcelulosa de forma individual o realizando *pools* (de 4-6 colonias). Las mismas fueron colocadas en 500µl de medio RPMI. Todos los tipos de colonias fueron utilizados para la extracción de ARN. Luego de una centrifugación, se resuspendieron los pellets en 100µl de solución de lisis (RNAqueous-Micro Kit, Ambion), y se conservaron a -20°C hasta el momento de extracción de ARN.

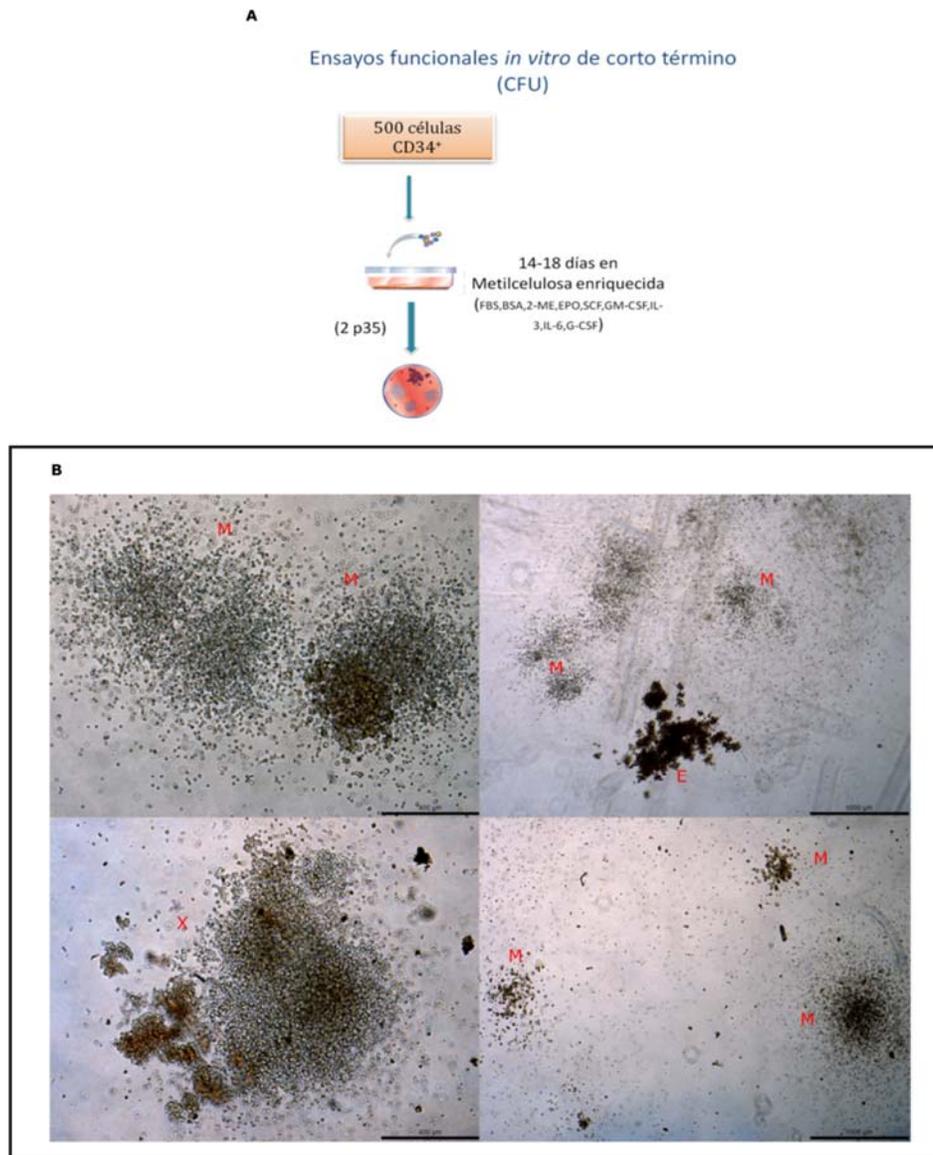


Fig. 2.1. A: Esquema del protocolo utilizado para los ensayos funcionales *in vitro* de corto término para células progenitoras (CFU). B: Fotografías de colonias representativas obtenidas en los ensayos CFU y LTC-IC. M: progenitores macrófago-granulocito (CFU-G/CFU-GM/CFU-M); E: progenitores eritroides (CFU-E/BFU-E); X: progenitores de linaje múltiple (CFU-GEMM).

#### II.4. Ensayo de largo término para células primitivas (LTC-IC).

El ensayo fue realizado en la modalidad *bulk* LTC (Fig. 2.2): se plaquearon  $\geq 5.000$  células CD34<sup>+</sup> (por duplicado en los casos donde fue suficiente la cantidad de células de la muestra) en un volumen de 2mL de medio de cultivo de largo término (LTCM) (Myelocult suplementado con hidrocortisona  $10^{-6}$ M, Stem Cell Technologies). Se utilizaron placas de cultivo p35 a las cuales se les realizó un pretratamiento con colágeno (BD Biosciences) para favorecer la adhesión de las *feeder layers*. Éstas consistieron en una mezcla 1:1 de líneas celulares de fibroblastos murinos (M2-10B4 IL-3, G-CSF y SI/SI IL-3, SCF), productoras de hIL-3, hSCF y hG-CSF (donadas por la empresa Stem Cell Technologies), que fueron irradiadas a 80Gy previo a la siembra de la muestra. Una vez por semana se cambió la mitad del medio por uno recién preparado. A las 6 semanas, se cosecharon las células adherentes y no adherentes. Para ello, las células se recolectan en un tubo, al que se le van agregando las siguientes fracciones: luego de un lavado con PBS, se levantaron las células adherentes con tripsina, seguido de un lavado adicional de la placa con IMDM/RPMI; las células se centrifugaron 7min a 1200rpm, se lavaron con 4mL de IMDM/RPMI suplementado con 2% de Suero Fetal Bovino (SFB) (Natocor, Córdoba, Argentina), y se resuspendieron en IMDM/RPMI 2% SFB para contabilizar células viables en cámara de Neubauer. Se plaquearon 20.000 células en metilcelulosa enriquecida (Methocult H4435) para la realización del ensayo CFU (por duplicado). Al cabo de 14-18 días de cultivo, se contaron las colonias al microscopio óptico, y se cosecharon las colonias individuales o *pooles* de colonias para extracción de ARNm (como se describió en la sección II.3)

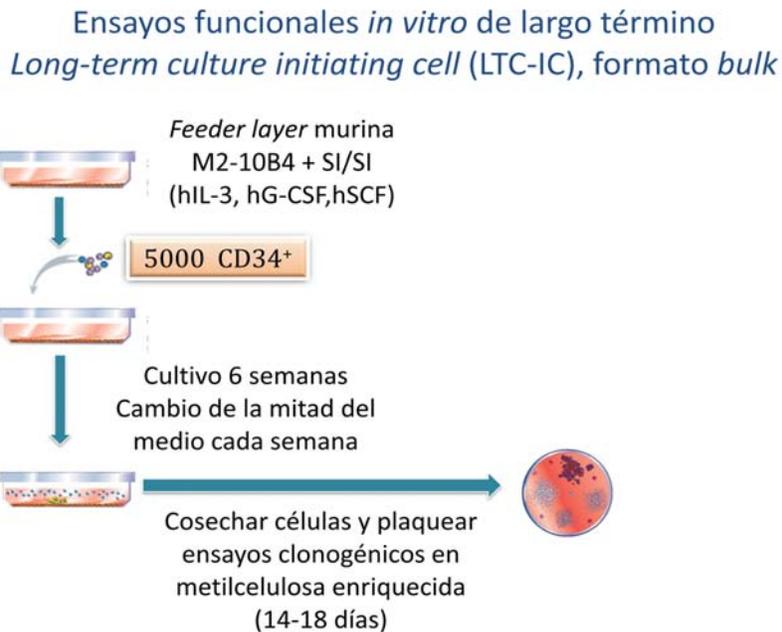


Fig. 2.2. Esquema del protocolo utilizado para los ensayos funcionales de largo término para células primitivas.

#### II.4.1 Cultivo de las líneas M2-10B4 y SI/SI.

Las células M2-10B4 IL-3, G-CSF fueron mantenidas en cultivo en botellas T75 en medio RPMI suplementado con suero SFB 10%, a una densidad inicial de  $2 \times 10^5$  células por botella; las células SI/SI SCF, IL-3 fueron mantenidas en cultivo en botellas T75 en medio DMEM alta glucosa (4500mg/L) suplementado con SFB 10%, a una densidad inicial de  $4 \times 10^5$  células por botella. Para cada ensayo LTC-IC se descongelaron ambas líneas (viales con menos de 10 pasajes), se seleccionaron con higromicina B (0,06 mg/mL para M2-10B4 y 0,125 mg/mL para SI/SI) y G418 (0,4 mg/mL para M2-10B4 y 0,8 mg/mL para SI/SI), y se irradiaron  $4 \times 10^6$  células de cada línea en 2,7mL de medio LTCM cada una, a 80Gy (la irradiación de las células fue realizada por el servicio de radioterapia del Instituto Alexander Fleming, CABA, Argentina). Luego se combinaron ambas líneas en proporción 1:1 y se plaquearon  $3 \times 10^5$  células totales por placa p35, en 2mL de LTCM. Los cultivos se iniciaron entre 1 y 7 días post-irradiación. La congelación de las líneas fue realizada en criotubos de 1,8mL conteniendo  $2-4 \times 10^6$  células, en SFB conteniendo 10% DMSO.

*Protocolo de coating con colágeno de las placas para la adhesión de las feeder layers:* diluir la solución de colágeno (*Collagen type I Rat Tail*, Cat. 354236, BD Biosciences) en ácido acético 0,02N ( $C_f=50\mu\text{g/mL}$ ); agregar 1-2mL a cada placa p35 ( $5\mu\text{g/cm}^2$ ); incubar 1h a TA; aspirar la solución remanente; enjuagar con 1mL de PBS; usar inmediatamente o secar al aire. Se pueden guardar a  $2-8^\circ\text{C}$  hasta 1 semana.

#### II.5. Extracción de ARN y RT-qPCR para la detección de transcritos *BCR-ABL1*.

Se extrajo ARN total a partir de colonias individuales o pools de colonias utilizando el kit RNAqueous-Microkit (Ambion), siguiendo el protocolo del fabricante. El ARN fue eluido en  $18\mu\text{L}$  de solución de elución a  $75^\circ\text{C}$ ; sólo se utilizaron aquellas muestras con pureza aceptable ( $A_{260}/A_{280} \geq 1,7$ ). Para la cuantificación del ARNm de *BCR-ABL1* en SP, se lisaron los rojos mediante 3 lavados en buffer amonio ( $144\text{mM NH}_4\text{Cl}$ ,  $10\text{mM NH}_4\text{HCO}_3$ ) de 15min cada uno, a TA con agitación suave. El ARN se extrajo con Trizol (Invitrogen), utilizando el protocolo de recomendado por el fabricante, con las siguientes modificaciones: las células fueron resuspendidas en Trizol y se dejó reposar a TA al menos 10min. En caso de no ser procesadas en el momento, las muestras se almacenaron a  $-80^\circ\text{C}$  hasta 1 mes. Se utilizaron  $200\mu\text{L}$  de cloroformo por cada 1mL de Trizol. Luego de vortexear se dejó reposar 10-15min a TA. Las centrifugaciones fueron realizadas a  $4^\circ\text{C}$ . La fase acuosa fue recuperada a un tubo nuevo, al que se agregó un volumen equivalente de isopropanol. Luego de mezclar por inversión 5 veces, se dejó reposar 10min a TA. Luego de la precipitación del ARN se agregó 1mL de EtOH 75% preparado en el día, se mezcló por inversión y se guardó a  $-20^\circ\text{C}$  por 24hs para favorecer la precipitación. Al día siguiente se centrifugó en frío 2 veces, y se eliminaron todos los restos de EtOH con micropipeta. La resuspensión del ARN fue realizada en  $\text{H}_2\text{O}$  apirógena a TA. La calidad del material obtenida fue evaluada por los cocientes  $A_{260}/A_{280}$  y  $A_{260}/A_{230}$  medidos en Nanodrop (Thermo Fisher Scientific).

El ARN fue sometido a retrotranscripción con Superscript II (Invitrogen) y *random primers* ( $C_f = 20\text{ng}/\mu\text{L}$ , Invitrogen), en volumen final de  $25\mu\text{L}$ , según el protocolo del fabricante. El ADNc fue diluido al medio con agua al finalizar la reacción de retrotranscripción, y almacenado a  $-20^\circ\text{C}$ . La qPCR fue realizada con *primers* (Invitrogen) ( $C_f=300\text{nM c/u}$ ) y

sondas (SIGMA Aldrich) (200nM) para *BCR-ABL1* y *ABL1* recomendados por el programa *Europe against cancer* <sup>9,68</sup>. La secuencia de los mismos está detallada en el Anexo II. Se utilizó el reactivo “Taqman Universal MasterMix II, no UNG” (Applied Biosystems). Se realizaron curvas de calibración para la cuantificación absoluta y estimación de la eficiencia de reacción mediante plásmidos conteniendo la secuencia del ADNc de *BCR-ABL1* y *ABL1* (BCR-ABL pDNA calibrant; ERM certified reference material ERMAD623, SIGMA).

*Protocolo de qPCR (Taqman) para BCR-ABL1 y ABL1:*

	1x (μL)	Programa	
<b>MasterMix 2X</b>	12,5	50°C 2min	
<b>Primer Fw 10μM</b>	0,75	95°C 10min	
<b>Primer Rv 10μM</b>	0,75	95°C 15seg	x 50 ciclos
<b>Sonda 10μM</b>	0,5	60°C 1min	
<b>H<sub>2</sub>O</b>	5,5		
<b>ADNc</b>	5		
<b>Vol. Final</b>	25		

Los resultados fueron armonizados a la escala internacional mediante un factor de conversión calculado a partir de la estandarización con calibradores secundarios<sup>27</sup>. Algunos pacientes fueron monitoreados en otros centros, en ese caso los cocientes %*BCR-ABL1<sup>IS</sup>* fueron obtenidos directamente de sus registros. En el caso de muestras al diagnóstico, se realizó una PCR cualitativa a partir del ADNc. Se utilizó la enzima GoTaq Flexi DNA polymerase (Promega).

*Protocolo de PCR cualitativa para BCR-ABL1 y ABL1:*

	1x p210 (μL)	1x ABL1 (μL)	Programa	
<b>Buffer Flexi 5X</b>	4	4	94°C 5min	
<b>Primer Fw 10μM</b>	1	1	94°C 30seg	x 50 ciclos
<b>Primer Rv 10μM</b>	1	1	60°C 30seg	
<b>dNTPs 10mM c/u</b>	1	1	72°C 30seg	
<b>MgCl<sub>2</sub> 25mM</b>	1,6	1,2	72°C 5min	
<b>H<sub>2</sub>O</b>	6,2	6,3		
<b>Taq pol.5U/μL</b>	0,2	0,2		
<b>ADNc</b>	5			
<b>Vol. Final</b>	25			

## II.6. Secuenciación masiva de la fracción de ARNs pequeños

### II.6.1. Citometría de flujo y *cell sorting*.

Se marcaron células CD34<sup>+</sup> o MNC con los siguientes anticuerpos (las cantidades indicadas son por cada 1x10<sup>6</sup> células marcadas): 5μL CD45-PerCP (2D1, BD Biosciences), 2,5μL CD34-FITC (AC136, Miltenyi Biotech), 2,5μL CD38-PE (IB6, Miltenyi Biotech), y 15μL CD26-APC (FR10-11G9, Miltenyi Biotech) en un volumen final acorde a la cantidad de células marcadas (50μL cada 500.000 células) de buffer MACS, durante 15 minutos a temperatura ambiente (TA). Las células fueron lavadas en 1mL de PBS y resuspendidas en

≥300 μL de PBS. Se utilizaron los citómetros de flujo FACS Canto II o FACS Aria II (BD Biosciences). En cada experimento se incluyó un control de isotipo, consistente en células marcadas con 2,5μL Mouse IgG2a-FITC, 2,5μL Mouse IgG2a-PE y 15μL Mouse IgG2a-FITC (todos de Miltenyi Biotech) y con 5μL CD45-PerCP, tal de establecer los cuadrantes sobre la población CD45<sup>low</sup>. En los casos de separación (*sorting*), se estableció un valor de *window extension* = 0 para evitar la contaminación con abortos electrónicos, se utilizó el nozzle de 70μm y el parámetro "pureza". Las células fueron recolectadas en condiciones asépticas, directamente en 100μL de buffer de lisis para extracción de ARN, en tubos de 200μL libres de RNAsa, o en 100μL de metilcelulosa enriquecida (Methocult H4435 Medium, Stem Cell Technologies) para la evaluación de pureza. Los datos de citometría de flujo fueron analizados con el programa FlowJo.

El ARN conteniendo la fracción de ARNs pequeños fue obtenido mediante el protocolo modificado del kit RNAaqueous (Ambion): al lisado se agregaron 125μL de EtOH 100%; todo el contenido se traspasó a una columna, colocada en un tubo recolector. Se centrifugó a máxima velocidad 1min y se realizó 1 lavado con la solución de lavado "1" (180 μL), y 2 lavados con solución de lavado "2/3" (180 μL c/u). Se descartó el eluido, y se centrifugó 1min a máxima velocidad para eliminar el líquido remanente. Luego la columna se insertó en un tubo recolector nuevo, libre de ARNasas, y se eluyó dos veces con 9 μL de H<sub>2</sub>O apirógena, precalentada a 75°C. El ARN fue guardado a -80°C hasta el momento de ser utilizado. En el caso de evaluar la pureza de las fracciones sorteadas, 500 células fueron recolectadas directamente en 250μL de Methocult, mezcladas inmediatamente y transportadas en hielo hasta el laboratorio. Allí fueron pasadas a una placa p35 con el agregado de 750μL de Methocult. El ensayo de cuantificación de colonias y aislamiento de ARN de estas fue realizado como se describió en el punto II.3. En los experimentos en los que se requirió la evaluación de la calidad del ARN por bioanalyzer, se contrató el servicio prestado por la Fundación Instituto Leloir.

#### **II.6.2. Concentración del ARN proveniente de distintas muestras para la secuenciación masiva del miRnoma.**

Las muestras fueron combinadas y congeladas a -80°C para ser transportadas en hielo seco, donde se almacenaron a -80°C hasta el momento de ser utilizadas. Se descongelaron en hielo y se concentraron por centrifugación con vacío, en tandas de 20min, cuidando de colocarlas en hielo de forma intermitente, para minimizar la degradación del ARN. Para evitar la degradación del material por ARNasas, todas las partes accesibles de la centrífuga fueron desinfectadas con EtOH 70°. Se resuspendieron en 7μL de agua apirógena estéril, incubando 5min a 50°C. Luego se mezcló repetidas veces, suavemente, para resuspender el material. Los pools utilizados para la construcción de la biblioteca fueron cuantificados por Qubit 2.0. En el caso de las líneas celulares, se utilizó 1μg de ARN para la construcción de la biblioteca, mientras que para las muestras de pacientes y dadores sanos se utilizó todo el volumen disponible (<140ng para los pools CD38<sup>-</sup> o CD38<sup>-dim</sup>, y 210ng para el pool CD38<sup>+</sup>).

### II.6.3 Construcción de bibliotecas de ARNs pequeños para secuenciación masiva en HiSeq 2500 (Illumina).

Se siguió el protocolo recomendado por el fabricante (Truseq Small RNA, Illumina) (Fig. 2.3). El mismo se basa en la ligación selectiva de ARNs con extremos 3'OH y 5'P libres, resultantes del procesamiento (clivaje) de precursores durante la biogénesis de ARNs pequeños. Esta estrategia también incluye a otros ARNs pequeños, por ej. fragmentos de ARNt, ARNs pequeños nucleolares (snoARNs), ARNs pequeños nucleares (snARNs); y piwiARNs (24-31nt).

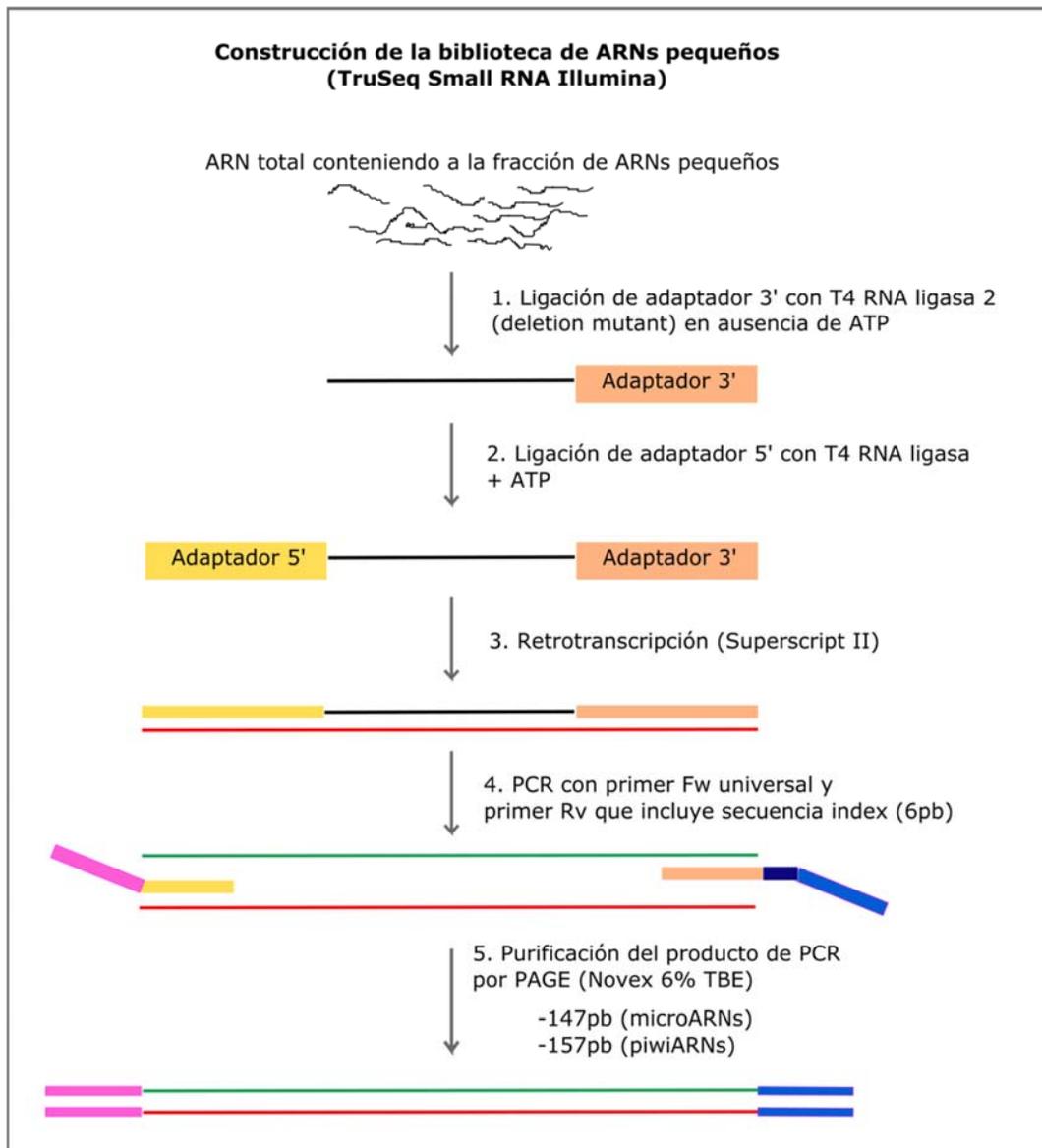


Fig. 2.3. Esquema de los primeros pasos realizados para la construcción de la biblioteca de ARNs pequeños.

1. Ligación de adaptadores adenilados al extremo 3' (usando una T4 RNA ligasa 2, mutante por delección, que en ausencia de ATP impide la ligación de ARNs pequeños entre sí); seguido del agregado de STOP solution;
2. Ligación de adaptadores al extremo 5' con T4 ARN ligasa en presencia de ATP;
3. Retrotranscripción con *primers* específicos que permiten enriquecer el producto de reacción en fragmentos con adaptadores en ambos extremos (Superscript RT II);
4. PCR que combina un *primer* Rv universal con un *primer* Fw conteniendo una secuencia index de 6pb distinta a cada muestra, lo que permitió la multiplexación durante la secuenciación (15 ciclos);
5. Purificación del producto de la biblioteca por electroforesis en gel de poliacrilamida 6% (Novex TBE PAGE Gel): se sembró el total del producto de reacción de PCR (1 muestra por calle). Luego de escindir con bisturí la región del gel comprendida entre los marcadores de peso molecular de 145 y 160 pb, se pasó el fragmento de gel a través de un tubo perforado, por centrifugación a 20.000xg 2min TA, y se dejó en agitación O.N. a TA para liberar el ADN del gel. Luego la solución fue pasada a través de un filtro de 5µm.
6. Cuantificación de las bibliotecas por qPCR (KAPA SYBR) utilizando *primers* complementarios a los adaptadores de Illumina, y una curva estándar. Las bibliotecas fueron almacenadas a -80°C hasta el momento de la corrida de secuenciación.
7. Concentración por centrifugación con vacío, y resuspensión en 1/3 del volumen original para la generación de *clusters* en la flow cell.
8. Secuenciación: *single-end*. Se utilizó la plataforma Hi-Seq 2500 (Illumina).

#### **II.6.4. Análisis bioinformático de los datos de secuenciación masiva del miRnoma.**

Se filtraron las lecturas de baja calidad y las secuencias contaminantes (adaptadores 3', 5', índices). La identificación de los microARNs conocidos fue realizada con la herramienta Chimira<sup>69</sup>. El cálculo de expresión diferencial se realizó con GFOLD, con  $c=0,01$  y mapeando contra una base de datos de snoRNA/miRNA (HISAT2)<sup>70</sup>. La búsqueda de microARNs *novels* fue realizada con miRDeep2<sup>71</sup>. Se seleccionaron aquellos con score  $\geq 1$ . El análisis *in silico* de los blancos y vías asociadas fue realizado con miRPath (*Diana tools*)<sup>72</sup> y ChemiRs<sup>73</sup>. Los parámetros utilizados para la búsqueda en miRPath fueron: "KEGG analysis", Tarbase (base de datos con interacciones experimentalmente validadas) o microT-CDS en el caso de que no existiesen evidencias experimentales, "Pathway union", " $p$ -value threshold"=0,001, "MicroT threshold"=0,8, "Enrichment analysis method= Fisher's exact test (hypergeometric distribution)", "FDR correction" (Benjamini & Hochberg), "Conservative stats". El alineamiento múltiple de la secuencia del novel-3 y las de miR-1839 encontradas en otras especies fue realizado con CLUSTAL OMEGA. Las intersecciones entre listados de microARNs fueron realizadas con el software R (v.3.4.0), mediante la función *intersect*. Se consultaron las bases de datos miRIAD<sup>74</sup>, miRBase y Gene (NCBI) para la caracterización genómica de los microARNs.

## II.7. Detección del punto de ruptura *BCR-ABL1* por secuenciación masiva.

### II.7.1 Preparación de muestras para la generación de bibliotecas

Se extrajo ADN de alta calidad (DNAzol, Thermo Fisher Scientific) a partir de muestras de SP o MO al diagnóstico. Se realizó una PCR *multiplex long-range* con  $\geq 500$ ng de ADN genómico, utilizando 4 sets de primers Rv y un *primer Fw* hibridando en el exón 13 o 14, según el rearrreglo *BCR-ABL1*, con modificaciones a partir del método publicado por Ross et al <sup>75</sup>. Las secuencias de los *primers* se encuentran en el Anexo II. Las condiciones de PCR para cada reacción multiplex fueron:

	1x Set A, B, D	1x Set C	[final]
Agua ( $\mu$ L)	4,075 $\mu$ L	5,075 $\mu$ L	25 $\mu$ L
Buffer sin $MgCl_2$	5 $\mu$ L	5 $\mu$ L	-
PCR	1,25 $\mu$ L	1,25 $\mu$ L	-
NucleotideMix			
Primers 10 $\mu$ M	1 $\mu$ L c/u	1 $\mu$ L c/u	400 nM
DMSO	0,375 $\mu$ L	0,375 $\mu$ L	1,5%
Enzima 5U/ $\mu$ L	0,3 $\mu$ L	0,3 $\mu$ L	1,5 U
$MgCl_2$ 25 mM	3 $\mu$ L	3 $\mu$ L	3 mM
ADNg (100ng/ $\mu$ L)	5 $\mu$ L	5 $\mu$ L	500 ng

La enzima utilizada para la PCR *long range* fue la *Expand<sup>TM</sup> long range dNTP pack* (Roche). El ciclado utilizado (de aproximadamente 7hs de duración) consistió en:

95°C	2min	10 ciclos
94°C	20seg	
66°C	30seg	
68°C	10min	
94°C	20seg	25 ciclos
66°C	40seg	
68°C	10min +10seg/ciclo	
68°C	9min	1 ciclo

El producto de PCR fue purificado a partir de un gel de agarosa 0,6%. El producto de la purificación fue cuantificado y utilizado para la generación de bibliotecas, según las instrucciones del fabricante. *En Ion torrent PGM (Thermo Fisher Scientific)*: fragmentación con Ion Shear Plus Reagents, purificación con Agencourt AMPure XP (Beckman Coulter), agregado de adaptadores e índices con Ion Xpress Plus Fragment Library Kit. Las bibliotecas fueron cuantificadas por qPCR. Finalmente, las muestras fueron preparadas para PCR en emulsión, enriquecidas con Ion PGM Template OT2 200 Kit, y secuenciadas en un Ion 318<sup>TM</sup> Chip v2 con Ion PGM Sequencing 200 Kit v2. *En MiSeq (Illumina)*: tagmentación (Amplicon tagment mix), amplificación y unión de adaptadores e índices (Nextera PCR Master mix), purificación de la biblioteca de tamaño óptimo (AMPure XP beads), normalización y pooleado de las muestras. Finalmente, las bibliotecas fueron secuenciadas en MiSeq (Illumina).

## II.7.2. Análisis bioinformático de los datos de secuenciación masiva para la identificación del punto de fusión *BCR-ABL1*.

### a) Datos de Ion Torrent PGM (Thermo Fisher Scientific)

Las lecturas fueron alineadas usando el *Subjunc aligner*, incluido en el paquete *Subread* (v1.5.0-p3) <sup>76</sup> y el genoma humano de referencia (versión hg19, UCSC). Los procesos de sorteado e indexado fueron realizados con Picard (v1.119) (<http://broadinstitute.github.io/picard>). Los puntos de fusión fueron localizados con CREST (Clipping Reveals Structure, v0.0.1)<sup>77</sup> y los resultados se verificaron por inspección visual de las lecturas alineadas a la región de la translocación (Integrative Genomics Viewer, IGV) (Fig. 2.4).

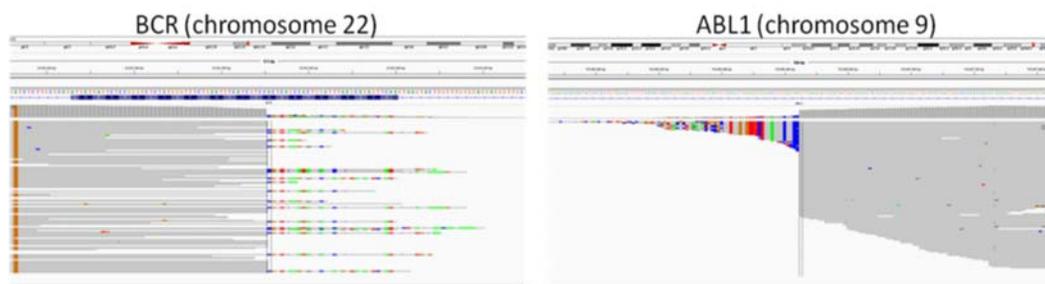


Fig. 2.4. Visualización de las lecturas provenientes de la secuenciación masiva del amplicón correspondiente al punto de fusión *BCR-ABL1* para la muestra N22. Cada fila es una lectura alineada al genoma de referencia. Se observa el punto de fusión como una zona de disminución abrupta en la cantidad de lecturas.

### b) Datos de MiSeq (Illumina)

Se realizó un filtrado de las lecturas de baja calidad y de contaminantes (Fastx-Toolkit, FastQC), seguido de pareado de las lecturas *paired-end* (Phyton script), mapeo y procesamiento de lecturas (BWA, GATK, Picard, Samtools) usando el genoma de referencia GRCh38. Los sitios de ruptura fueron detectados mediante el algoritmo Breakdancer<sup>78</sup>, y los resultados se verificaron por inspección visual de las lecturas alineadas a la región de la translocación (IGV).

Finalmente, tanto para los datos de Ion Torrent como de MiSeq, se usó el algoritmo Blast2Seq contra las secuencias de referencia *ABL1* y *BCR* para poder identificar cada segmento, y se diseñaron primers específicos con el software Primer3plus<sup>79</sup> (sintetizados por Genbiotech o Invitrogen).

## II.7.3. Detección de células *BCR-ABL1*<sup>+</sup> en la fracción primitiva a nivel de ADN genómico.

La secuencia correspondiente al punto de fusión paciente-específico se utilizó para el diseño de primers para PCR. Los cultivos para células primitivas y progenitoras fueron realizados como se describe en las secciones II.4 y II.5. Se extrajeron colonias individuales de la metilcelulosa (con micropipeta p20 fijada en 5µL, utilizando tips con filtro) y se colocaron en 50 µL de: buffer GoTaqFlexi PCR (Promega, USA), NP-40 0,5%, Tween-20 0,5%, y proteinasa K 0,91 mg/mL, incubadas 1h a 60°C + 15 minutos a 95°C. Se usó un

10% del volumen final para la primera ronda de PCR. El producto de ésta se diluyó 10 a 1000 veces para la segunda ronda de PCR. Los productos fueron evaluados por electroforesis en gel de agarosa 2%. En aquellos casos en los que el peso molecular del producto no se alineaba exactamente con el control positivo, se secuenció el producto de PCR por Sanger (Macrogen, Estados Unidos).

	N35 I y II	MP5920 I y II	N38 I	N38 II	N22 I y II
<b>dNTPs</b>	0,2mM	0,2mM	0,2mM	0,2mM	0,2 mM
<b>Primer Fw</b>	0,5 µM	0,5 µM	0,5 µM	0,3µM	0,5 µM
<b>Prime Rv</b>	0,5 µM	0,5 µM	0,5 µM.	0,3 µM	0,5 µM
<b>MgCl<sub>2</sub></b>	2mM	1,5mM	1,5mM	1,2mM	1,5 mM
<b>TaqPol</b>	0,0625 U/µL	0,025 U/µL	0,025 U/µL	0,025 U/µL	0,025 U/µL
<b>Temp. ann.</b>	60°C	60	60	63	59°C/57°C
<b>No. ciclos</b>	35	35	35	35	35

	B-RAF (cualitativa)	B-RAF (qPCR)
<b>dNTPs</b>	0,5 mM	<b>PowerUp SYBR Green Mastermix (ABI)*</b> 1X
<b>Primer Fw</b>	0,5 µM	<b>Primer Fw</b> 800nM
<b>Prime Rv</b>	0,5 µM	<b>Primer Rv</b> 800nM
<b>MgCl<sub>2</sub></b>	2 mM	
<b>TaqPol</b>	0,0625 U/µL	
<b>Temp. ann./No. ciclos</b>	57°C/ 35	<b>Programa</b> 50°C 2min + 95°C 2min + (95°C 15seg + 57°C 1min) x40

\*Según los datos del fabricante, la concentración de Mg<sup>2+</sup> en la mix puede variar entre 4,76-6,44mM.

## II.8. Cuantificación de la expresión de microARNs por stem-loop RT-qPCR.

Se utilizó un protocolo modificado a partir del desarrollado por Chen *et al*, basado en la retrotranscripción (RT) utilizando *primers stem-loop* específicos<sup>80</sup>, seguida de reacciones individuales de qPCR en presencia del agente intercalante fluorescente SYBR Green, utilizando un *primer Fw* específico de microARN, en combinación con un *primer Rv* universal que hibrida con la región constante del primer *stem-loop* (Fig. 2.5). La RT fue realizada en 2 reacciones multiplex (M1 y M2). En paralelo se utilizó la misma cantidad de ARN para retrotranscribir el ARN pequeño snARN U6, utilizado como gen de referencia durante la cuantificación posterior. Las concentraciones finales en la reacción de RT (volumen final= 20µl) fueron: dNTPs 0,25mM (Invitrogen o INBIO Highway); DTT 10mM (Invitrogen); Superscript II 2,5U/µL (Invitrogen); RNASEout 0,2U/µL (Invitrogen); *stem-loop primer* 0,05µM (de cada uno) o random primers (para snARN U6) 0,01µg/µL (Invitrogen). El ARN fue incubado con H<sub>2</sub>O<sub>d</sub> y dNTPs durante 5min a 65°C, seguido de 1min en hielo; luego de realizar un spin, se agregaron los componentes restantes de reacción y se incubaron: 30min 16°C + (30seg 30°C + 30seg 42°C + 1seg 50°C) x 40 + 5min 85°C. El ADNc fue diluido al medio con H<sub>2</sub>O<sub>d</sub> y guardado a -20°C hasta su utilización. Para la qPCR, se utilizaron 2µL de ADNc diluido por reacción, en presencia de primer Fw 0,05µM; primer Rv universal 0,05µM y SYBR Green (PowerUp SYBR Green MasterMix, Applied Biosystems; según los datos del fabricante, la concentración de Mg<sup>2+</sup> en la mix puede variar entre 4,76-6,44mM). El programa de qPCR consistió en 2min 50°C + 2min

95°C + (15seg 95°C + 1min 60°C) x50. Se utilizó el equipo Rotor-Gene Q (Qiagen). Todas las determinaciones fueron seguidas por una curva de melting para evaluar la especificidad de la reacción (Fig. 2.6), y fueron realizadas por duplicado en corridas separadas. En aquellos casos donde los duplicados diferían en Ct>2 entre corridas, se realizaron triplicados. Los cálculos de eficiencia de amplificación fueron realizados mediante curvas de masa para la RT, mediante la fórmula  $[10^{(-1/m)}]-1$ , siendo  $m$  la pendiente de la curva de masa.

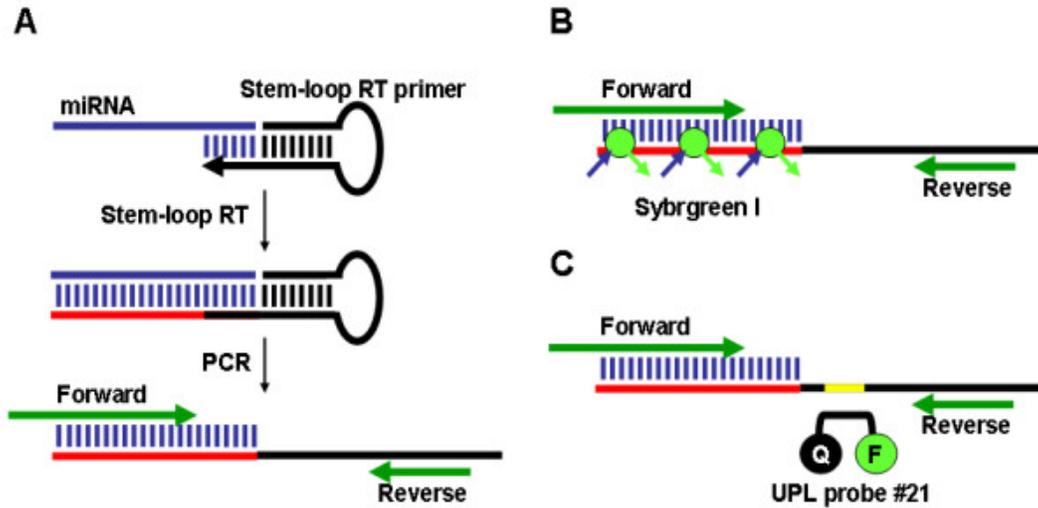


Fig. 2.5. Esquema de la estrategia de RT-qPCR de microARNs utilizando un primer stem-loop específico durante la retrotranscripción, y un primer Rv universal, complementario a la región del stem-loop, junto con un primer Fw específico de cada microARN, para la qPCR. La detección fue realizada por SYBR Green. Extraído de <sup>80</sup>

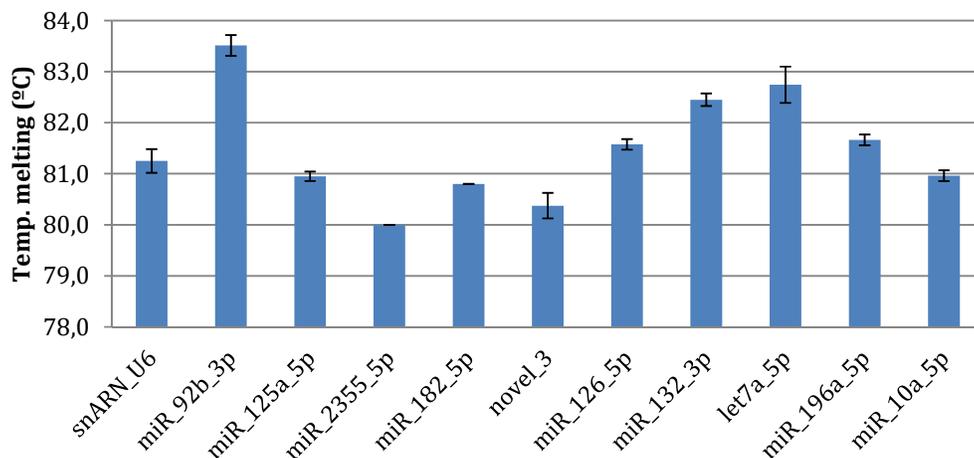


Fig. 2.6. Promedio de la temperatura de melting detectada para todos los microARNs evaluados, y para snARN U6, a partir de los datos provenientes de dos líneas celulares y una muestra de paciente

## II.9. Análisis estadístico y construcción de gráficos y figuras.

Se utilizaron los programas GraphPad Prism 6, Microsoft Excel 2007, Inkscape 0.92 y Benchling para la realización de gráficos, e Infostat v.2018e (Córdoba, Argentina) para el análisis estadístico. Cuando no se cumplieron los supuestos de normalidad y/o homocedasticidad de los datos, se modeló la varianza (función VarIdent) en caso de realizar un test paramétrico (ANOVA, modelos lineales mixtos), o se analizaron los datos con tests no paramétricos (Mann-Whitney  $U$  o Kruskal Wallis). En el caso de los datos de cuantificación de los niveles de microARNs por RT-qPCR, la variable analizada fue  $dCT = (Ct \text{ microARN } X - Ct \text{ snARN } U6)$ . Se ingresó la variable "Población" (LMC LSC, LMC HSC, LMC progenitores, DS HSC, DS Progenitores) como factor fijo, y la variable "Muestra", correspondiente a la identificación de cada paciente o dador sano, como factor de correlación (simetría compuesta). Se modeló la varianza con la función VarIdent, utilizando la variable "Población". Se controló el *false discovery rate* multiplicando el valor  $p$  obtenido en la prueba global por el número de variables analizadas sobre el mismo set de datos (8 microARNs). Las comparaciones *a posteriori* fueron realizadas con la fórmula DCG<sup>81</sup>.

## RESULTADOS

### Capítulo III. Evaluación de la carga leucémica como biomarcador de respuesta al tratamiento con ITK en la fracción primitiva y progenitora de pacientes con LMC

#### III.1. Introducción

El sistema hematopoyético es un continuo de células con fenotipo y propiedades que varían desde las más primitivas hasta los estadios diferenciados. Dada la insuficiencia de marcadores fenotípicos y la heterogeneidad funcional de las distintas poblaciones que componen el sistema hematopoyético, la utilización de ensayos funcionales indirectos y retrospectivos, descritos por primera vez hace más de 30 años, continúa siendo esencial para su estudio.

Los ensayos de cultivo de largo término de células iniciadoras (LTC-IC) y de corto término de unidades formadoras de colonias (CFU) son ensayos funcionales *in vitro* que permiten, de forma retrospectiva, cuantificar células de la fracción primitiva y progenitora comisionada a linaje, respectivamente. Ambos miden dos parámetros de importancia para el estudio de estas poblaciones: capacidad proliferativa (estimada por el número de células producidas) y potencial de diferenciación (estimado por el número de linajes diferentes representados en la progenie)<sup>82</sup>. El ensayo de LTC-IC se basa en la premisa de que las células más primitivas (que incluyen a las HSC) requieren de un gran número de divisiones celulares (>15, aproximadamente 5 semanas *in vitro*) antes de poder producir una progenie de células diferenciadas; en cambio, los progenitores comprometidos a linaje requieren de sólo 5-10 divisiones (2-3 semanas *in vitro*) para diferenciarse<sup>82</sup>.

Las condiciones que permiten el crecimiento de las fracciones de células primitivas y progenitoras también varían. Los factores de crecimiento requeridos por las células progenitoras comisionadas a linaje son conocidos y producidos como proteínas recombinantes; en cambio para la fracción de células primitivas aún no está estandarizada una única condición experimental, por lo tanto, existen varios protocolos para el ensayo LTC-IC.

Con el objetivo de evaluar si la dinámica de estas poblaciones varía en pacientes al diagnóstico y bajo tratamiento con ITK, se realizaron ambos ensayos en muestras de SP o MO de pacientes con LMC. Nuestra hipótesis es que el nivel de carga leucémica en la fracción primitiva es un indicador de la respuesta al tratamiento.

#### III.2. Pacientes incluidos en el estudio y puesta a punto del ensayo

Los pacientes incluidos en esta etapa del proyecto (N=38) fueron clasificados según se encontraban al diagnóstico, previo inicio del tratamiento con ITK, o en tratamiento con 1<sup>ra</sup> o 2<sup>da</sup> línea de ITK. A su vez, los pacientes en tratamiento fueron clasificados según su RM en SP (Tabla 3.1).

Tabla 3.1. Características clínicas de los pacientes incluidos en el estudio

	<b>Al diagnóstico</b>	<b>RM<sup>1.0</sup>-RM<sup>2.0</sup></b>	<b>RM<sup>3.0</sup>-RM<sup>5.0</sup></b>
	(N=12)	(N=13)	(N=13)
<b>Edad (años)</b>			
<b>Mediana</b>	49	62	67
<b>Rango</b>	18-79	36-74	28-75
<b>Género F/M</b>	5/7	2/11	6/7
<b>Score Sokal –no. casos</b>			
<b>Bajo</b>	4 *	5	5 *
<b>Intermedio</b>	5 *	4	5 *
<b>Alto</b>	2 *	4	2 *
<b>ITK –no. casos</b>			
<b>Imatinib</b>	n/a	13	9
<b>Nilotinib 1<sup>a</sup> línea</b>	n/a	0	3
<b>Nilotinib 2<sup>da</sup> línea</b>	n/a	0	1
<b>Tiempo de tratamiento con ITK (meses)</b>			
<b>Mediana</b>	n/a	6	18
<b>Rango</b>	n/a	3-9	6-85

Al diagnóstico: antes de iniciar el tratamiento con ITK; RM: Respuesta molecular en SP; n/a: no aplicable, \*:1 dato faltante

Los ensayos CFU para progenitores comprometidos a linaje fueron realizados en metilcelulosa enriquecida en factores de crecimiento. La utilización de un medio semisólido permitió cuantificar el número de progenitores iniciales en la muestra, y diferenciar el linaje de la progenie (ver Materiales y métodos).

Los ensayos LTC-IC para células primitivas fueron realizados con *feeder layers* murinas, secretoras de hIL-3 (*human Interleukin-3*), hSCF (*human Stem cell factor*) y hG-CSF (*human Granulocyte Colony Stimulating Factor*) (ver Materiales y métodos). La duración del ensayo fue de 6 semanas. En la bibliografía se encuentran reportes de este ensayo con duración variable, entre 5 y 8 semanas<sup>83</sup>; se seleccionó este tiempo para poder detectar una fracción primitiva en un tiempo adecuado para el trabajo experimental.

En primer lugar, comparamos la capacidad proliferativa, en ensayos de corto y largo término entre pacientes con distinta RM en SP. En los ensayos CFU la capacidad proliferativa se estimó a partir del número de colonias totales, mientras que en ensayos LTC-IC se tomó como estimador al aumento de células durante el cultivo de 6 semanas, ya que esta variable presentó menor coeficiente de variación entre réplicas provenientes del mismo paciente en comparación con el número total de colonias. En ensayos CFU la capacidad proliferativa fue similar entre los grupos ( $p > 0,05$ , ANOVA), mientras que en ensayos LTC-IC fue menor en las muestras al diagnóstico y en RM1.0 ( $p < 0,05$ , ANOVA, comparaciones *a posteriori*) (Fig. 3.1).

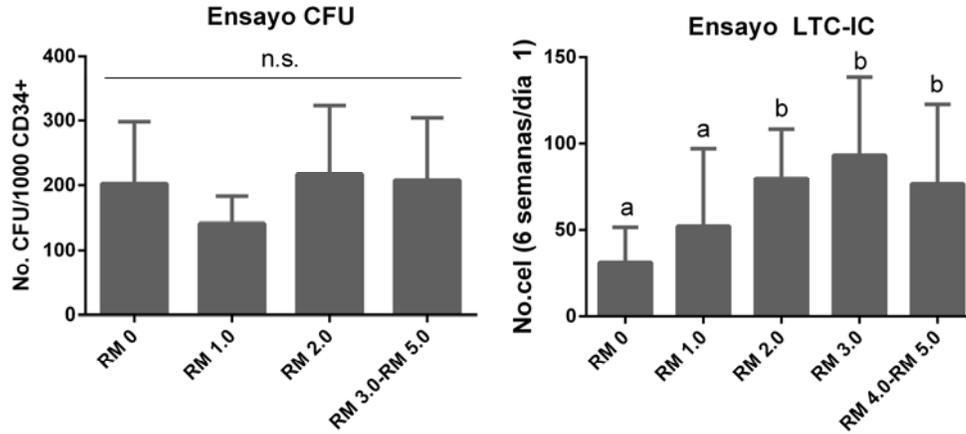


Fig. 3.1. Comparación en la capacidad proliferativa en ensayos CFU y LTC-IC de muestras de pacientes al diagnóstico (RM 0) o bajo tratamiento con ITK, con distinto nivel de RM en SP. Letras distintas indican diferencias estadísticamente significativas ( $p < 0,05$ , ANOVA, comparaciones a posteriori).

Por otro lado, se observó que la capacidad de diferenciación, estimada a través de la proporción de los distintos linajes en el ensayo clonogénico, fue diferente en ambos ensayos. Por un lado, en ensayos CFU, se observó una mayor proporción de progenitores eritroides (BFU-E/CFU-E) y una menor proporción de progenitores de tipo macrófago/granulocito (CFU-G/CFU-M/CFU-GM) en muestras de pacientes al diagnóstico en comparación con las provenientes de pacientes bajo tratamiento con ITK (Fig. 3.2, prueba Kruskal Wallis,  $p < 0,05$ ). En ensayos de largo término, la proporción de los distintos tipos de colonias fue similar para todos los grupos (Fig. 3.2, prueba Kruskal Wallis,  $p > 0,05$ ).

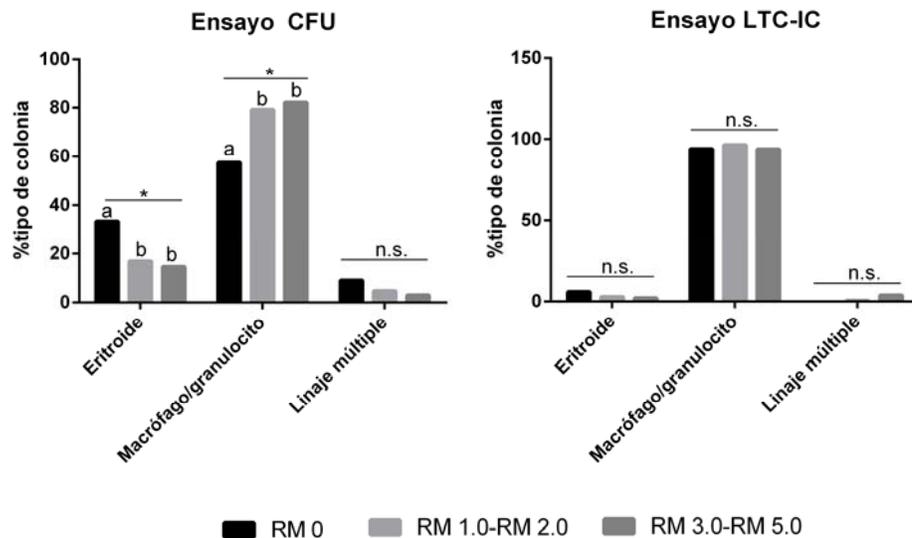


Fig. 3.2. Proporción de cada tipo de colonia en ensayos CFU y LTC-IC, en muestras de pacientes al diagnóstico (RM 0) o bajo tratamiento con ITK, con distinto nivel de RM en SP. Letras diferentes indican diferencias estadísticamente significativas ( $* = p < 0,05$ , prueba Kruskal Wallis).

Por otro lado, se observó una mayor proporción de colonias del linaje eritroide en ensayos CFU (media=22%, SD=13%) vs. LTC-IC (media=2%, SD=3%) (Fig. 3.3,  $p<0,05$ , prueba Mann-Whitney  $U$ ), lo que sugiere que las condiciones experimentales del cultivo de largo término desfavorecen el mantenimiento y proliferación de células de este linaje (Fig. 3.3).

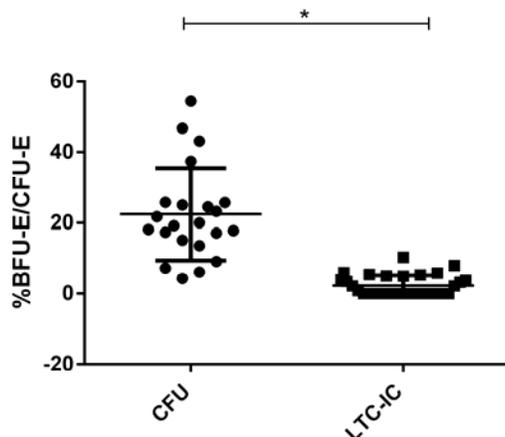


Fig. 3.3. Proporción de colonias del linaje eritroide en ensayos CFU y LTC-IC, combinando muestras al diagnóstico y bajo tratamiento con ITK.  $*=p<0,05$ , prueba Mann-Whitney  $U$ .

Con el objetivo de evaluar la proporción del clon leucémico en la fracción de células primitivas y progenitoras, se puso a punto la metodología de detección del ARNm de *BCR-ABL1* en las colonias provenientes de los ensayos de corto y largo término. En primer lugar, se evaluaron distintos métodos de extracción de ARN a partir de colonias individuales, resultando más adecuado un micrométodo basado en columna, que el método tradicional de separación de fases (Trizol), en cuanto a calidad del material obtenido (Fig. 3.4, A-C).

Utilizando una muestra de un paciente al diagnóstico, en la cual esperábamos una alta frecuencia de células leucémicas, observamos que las colonias individuales provenientes de ensayos LTC-IC expresaban menores niveles de ARNm *BCR-ABL1* ( $\%BCR-ABL1/ABL1$ : media  $\pm$  SD= 0,2% $\pm$ 0,3%, n=15) en comparación con aquellas provenientes de ensayos CFU ( $\%BCR-ABL1/ABL1$ : media  $\pm$  SD= 32% $\pm$ 24%, n=7), y una menor proporción de colonias *BCR-ABL1+* (33% vs. 100%). Por lo tanto, con el objetivo de aumentar la sensibilidad del ensayo, se realizó un ensayo LTC-IC, a partir del cual las colonias obtenidas fueron procesadas individualmente o en combinaciones (*pooles*) de 4-6 colonias previo a la extracción de ARN. El ensayo se realizó a partir de una muestra de MO de un paciente bajo tratamiento con ITK por 9 meses. En el caso de los *pooles* de colonias, se obtuvo un mayor rendimiento de ARN, y una mayor proporción de muestras con buena calidad, evaluada por la relación de absorbancia 260nm vs. 230nm (Fig. 3.4, D-F). Esto permitió aumentar la sensibilidad del ensayo, al evaluar un mayor número de colonias (a través de *pooles*) por paciente, y al obtener un mayor número de copias del gen control *ABL1* (Fig. 3.4, G). A su vez, una muestra de SP del mismo paciente fue utilizada para un ensayo LTC-IC, y las colonias fueron procesadas como *pooles*. Los resultados obtenidos fueron similares a lo observado para las colonias provenientes de MO (Fig. 3.4, D-G).

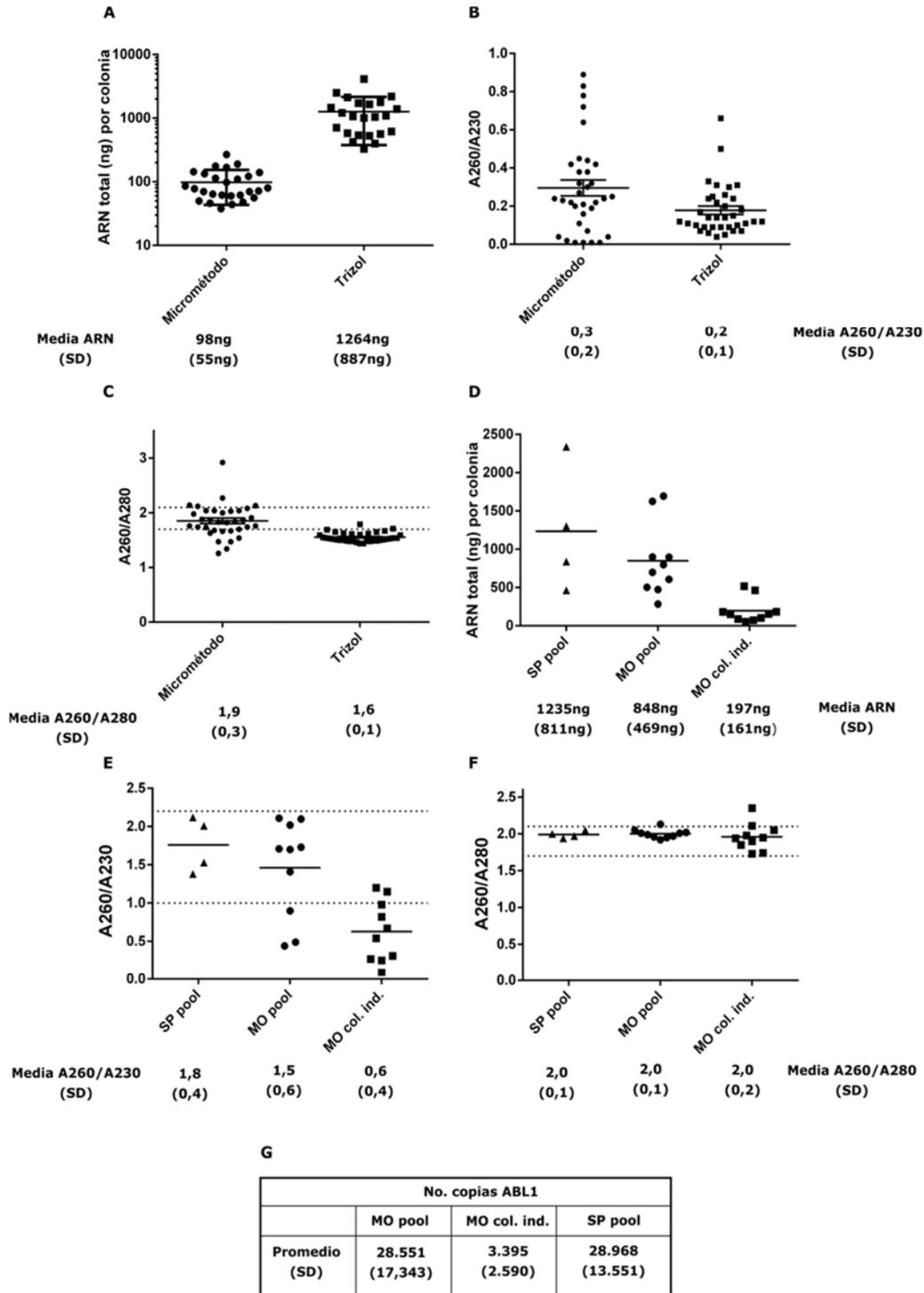


Fig 3.4. Comparación de rendimiento y calidad entre dos métodos de extracción de ARN a partir de colonias individuales (A, B, C). Comparación del rendimiento y calidad del ARN obtenido utilizando micrométodo a partir de colonias individuales y pools de colonias de MO, y a partir de pools de colonias de SP (D, E, F). El número de copias de ABL1 en pools de colonias de MO y SP fue en promedio 8,5 veces mayor al detectado en colonias individuales de MO (G).

### III.3. Evaluación de la carga leucémica en pacientes con LMC al diagnóstico y durante el tratamiento con ITK

Para comparar el nivel de carga leucémica en las fracciones primitivas y progenitoras entre los distintos grupos de pacientes, se calculó la mediana del cociente %*BCR-ABL1/ABL1* de las colonias individuales (sólo en el caso de los ensayos CFU al diagnóstico) o de los pools de colonias pertenecientes al mismo paciente. De ahora en más, y a lo largo del capítulo III, se simplificará la terminología y se hablará siempre de "colonias" en vez de "pools de colonias" o "colonias individuales". En pacientes al diagnóstico, previo tratamiento con ITK, los niveles de carga leucémica fueron significativamente menores en ensayos LTC-IC vs. CFU (media±SD: 6,3%±13,2% vs. 52,4%±21,5% respectivamente, prueba Mann-Whitney *U*,  $p < 0,05$ ). Por otro lado, la variabilidad en la carga leucémica entre pacientes fue mayor en la fracción primitiva que la progenitora, siendo que la primera abarcó un rango de más de 2-log (Fig. 3.5, A).

La carga leucémica en ambas fracciones fue sucesivamente menor en muestras provenientes de pacientes bajo tratamiento con ITK por 3, 6 o  $\geq 12$  meses (Fig. 3.5, B-D).

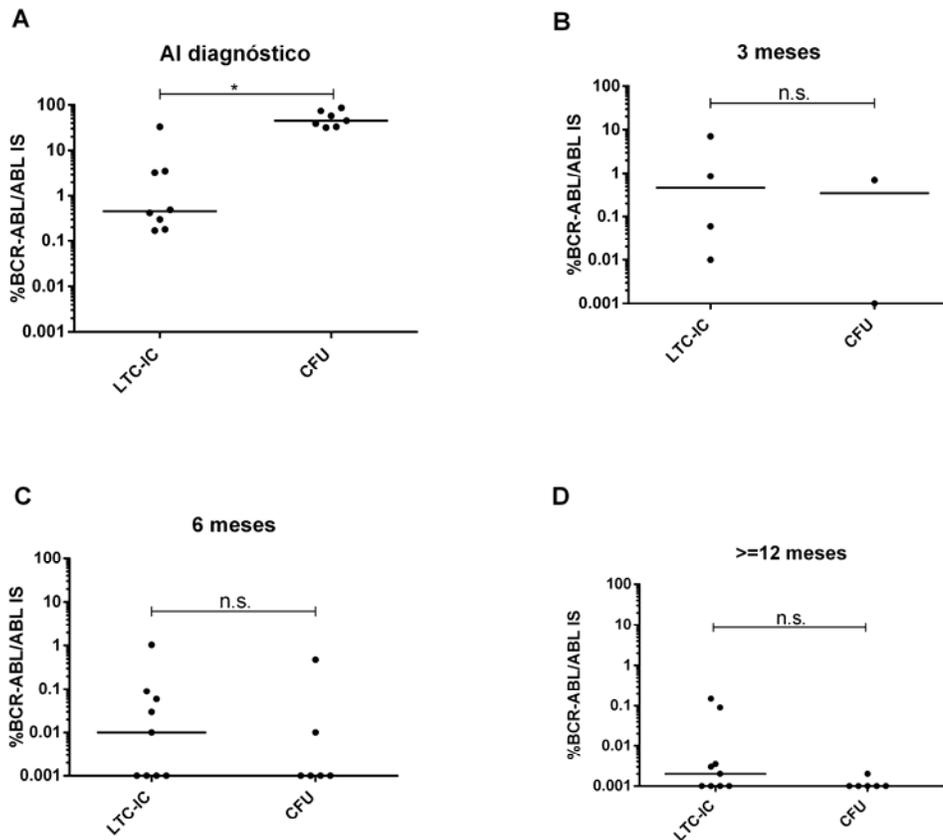


Fig 3.5. Niveles de carga leucémica en ensayos LTC-IC o CFU, al diagnóstico y durante el tratamiento con ITK. Cada punto corresponde a la mediana de %*BCR-ABL1/ABL1* de las colonias provenientes de un paciente individual. Los valores de mediana = 0 fueron asignados como 0,001%, correspondiente al límite teórico para detección de *BCR-ABL1/ABL1* por RT-qPCR. \* =  $p < 0,05$ ; prueba Mann-Whitney *U*.

Al considerar el porcentaje de colonias que expresan ARNm *BCR-ABL1*, independientemente de los niveles de expresión de cada una, se observó que, una vez iniciado el tratamiento, para el mismo nivel de RM en SP, los niveles de carga leucémica eran mayores en la fracción primitiva, en comparación con la progenitora (Fig. 3.6). Este patrón concuerda con las características de dicha fracción, cuyas células son poco proliferativas, de modo que requerirían más tiempo de exposición a ITK que las células progenitoras, para agotar el reservorio.

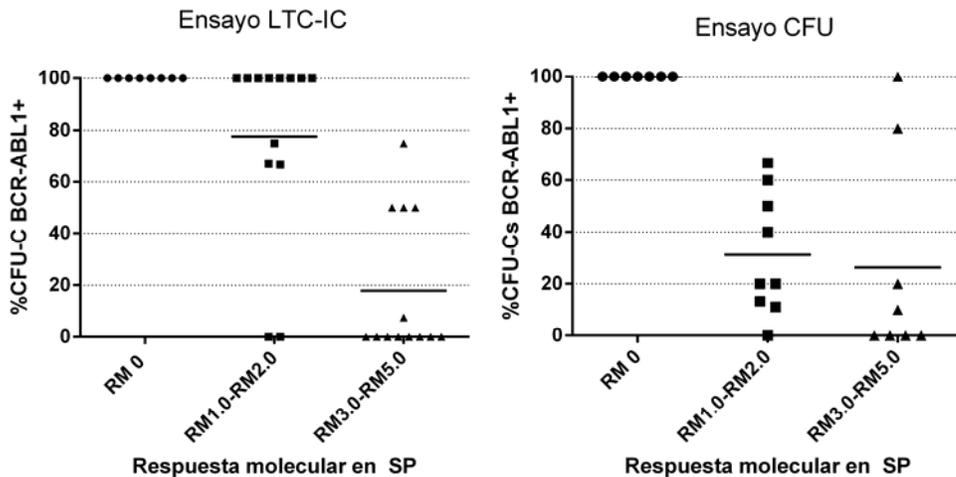


Fig. 3.6. Porcentaje de colonias *BCR-ABL1+* por paciente, en función del nivel de RM en SP. "RM 0" corresponde a muestras de pacientes al diagnóstico, sin tratamiento con ITK.

#### III.4. Comparación de la carga leucémica en pacientes con distinta respuesta clínica al tratamiento con ITK

Según las guías de tratamiento, la reducción temprana de la carga de transcritos *BCR-ABL1* en SP y/o del porcentaje de metafases con cromosoma Ph (Ph+) en MO, está asociada a una respuesta óptima al tratamiento con ITK en primera línea, mientras que aquellos que responden de manera más lenta son categorizados como "advertencia"; los pacientes que no muestran respuesta alguna desde el inicio son categorizados como "falla" al tratamiento (ver Introducción, tabla 1.1)<sup>24</sup>. Aunque la mayoría de los pacientes logran mantener una RM adecuada a lo largo del tiempo, se estima que a los 5 años de tratamiento con IM, hasta un 50% de ellos puede desarrollar resistencia, intolerancia o progresión de la enfermedad, requiriendo el cambio a un ITK de segunda línea, o trasplante de células precursoras hematopoyéticas en el caso de progresión<sup>17</sup>. En este contexto, es de interés la posibilidad de contar con marcadores pronósticos adicionales que puedan separar grupos con diferente riesgo de presentar una respuesta inadecuada al tratamiento. Bajo la hipótesis de que la carga leucémica en la fracción primitiva y/o progenitora puede ser un biomarcador de respuesta al tratamiento, comparamos los niveles de carga leucémica entre grupos de pacientes categorizados como de respuesta "óptima" o "advertencia/falla", hasta un seguimiento de 12 meses de tratamiento con ITK (tabla 3.2).

Tabla 3.2. Valores de los parámetros clínicos evaluados (hasta los 12 meses de tratamiento) para la clasificación de los pacientes en categorías de respuesta al tratamiento con ITK 1ª línea según la European Leukemia Net. Se listan sólo aquellos pacientes cuyas muestras fueron recibidas hasta los 12 meses de tratamiento (inclusive). RM = respuesta molecular; Ph+: porcentaje de metafases con presencia del cromosoma Ph. N/A= no aplicable; ND= no determinado

<b>Respuesta a ITK 1ª línea. Categoría "óptima"</b>						
Código paciente	RM (3 meses)	RM (6 meses)	RM (12 meses)	Rotación a ITK 2ª línea	Intolerancia a ITK	Progresión
N35	≤10%	<1%	≤0,1%	No	No	No
N42	≤10%	<1%	≤0,1%	No	No	No
N14	≤10%	<1%	≤0,1%	No	No	No
N23	≤10%	<1%	≤0,1%	No	No	No
N31	≤10%	<1%	≤0,1%	No	No	No
N10	≤10%	<1%	≤0,1%	No	No	No
N18	≤10%	<1%	≤0,1%	No	No	No
N14	≤10%	<1%	≤0,1%	No	No	No
LSC18	≤10%	<1%	≤0,1%	No	No	No
<b>Respuesta a ITK 1ª línea. Categoría "Advertencia"/"Falla"</b>						
	RM (3 meses)	RM (6 meses)	RM (12 meses)	Rotación a ITK 2ª línea	Intolerancia a ITK	Progresión
LSC10	>10%	≥1%	>0,1%	Sí (>12 meses)	Sí	No
N13	>10%	≥1%	N/A	Sí (<12 meses)	ND	ND
N28	>10% (25% Ph+)	≥1%	N/A	Sí (9 meses)	ND	No
N1	≤10% (25% Ph+ a los 2 meses)	N/A (0% Ph+ a los 5 meses)	N/A	Sí (9 meses)	Sí	No
N8	≤10%	≥1% (en el límite)	N/A	Sí (11 meses)	No	Sí (antes de la rotación)
N29	≤10% (100% Ph+)	<1% (10% Ph+)	ND	Sí (26 meses)	No	No
N22	≤10%	<1% (6% Ph+)	≤0,1%	No	No	No

Los datos de carga leucémica fueron analizados de dos maneras:

- a) considerando la mediana de %BCR-ABL1/ABL1 de las colonias de cada paciente;
- b) comparando los valores de %BCR-ABL1/ABL1 de cada colonia en cada grupo, sin diferenciar entre pacientes

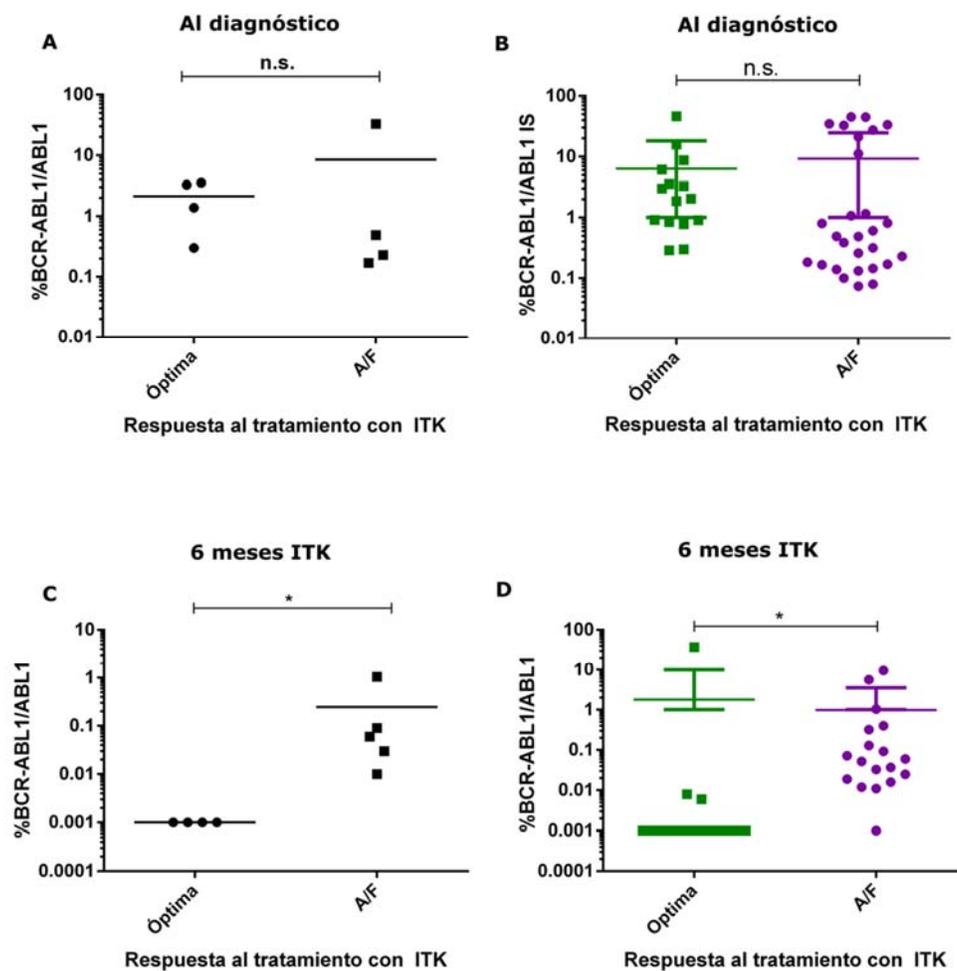


Fig. 3.7. Comparación de la carga leucémica en la fracción primitiva mediante ensayos LTC-IC con muestras provenientes de pacientes al diagnóstico, previo tratamiento (A-B) o a los 6 meses de tratamiento con ITK 1ª línea (C-D). \*= $p < 0,05$ ; A/F = "advertencia/falla"

Al diagnóstico no detectamos diferencias estadísticamente significativas en la fracción de células primitivas entre pacientes con respuesta óptima y advertencia/falla, con ambos métodos (Fig. 3.7, A:  $p > 0,05$ , prueba Mann-Whitney  $U$ ; B:  $p > 0,05$ , ANOVA).

A los 6 meses de tratamiento se detectaron diferencias estadísticamente significativas entre los grupos de respuesta, siendo más alta la carga leucémica en el grupo "advertencia/falla" ( $p < 0,05$ , prueba Mann-Whitney  $U$ ) (Fig. 3.7, C-D).

Aunque sería conveniente contar con más muestras en cada grupo, los resultados sugieren que, a los 6 meses de tratamiento, la carga leucémica en la fracción de células primitivas correlaciona con lo que ocurre a nivel sistémico, ya que en 4 de los 5 casos, los signos de alarma se evidenciaron a los 6 meses, ya sea en su RM en SP (N28), o en la respuesta citogenética en MO (N29, N22) o por signos de intolerancia al tratamiento (N1). Sólo el caso N8 resultó dudoso, ya que la RM a los 6 meses fue de 1,6%, muy cercana al límite (1%), y dada la variabilidad intrínseca del método (CV=17-30%, según el nivel de RM), es un dato no concluyente. A su vez, a los 8 meses de tratamiento, su RM había disminuido a 0,8%, sugiriendo que podría categorizarse como respuesta óptima. Sin embargo, en ese

momento presentó progresión (sarcoma granulocítico), que pudo ser tratado de manera exitosa mediante radioterapia.

Respecto a la carga leucémica en células progenitoras al diagnóstico, ésta fue mayor en el grupo de pacientes de respuesta "óptima" vs aquellos categorizados como de respuesta "advertencia/falla", considerando a las colonias individuales (Fig. 3.8, B;  $p < 0,05$ , ANOVA). Esta tendencia se observó analizando las medianas por paciente, aunque las diferencias no resultaron estadísticamente significativas (Fig. 3.8, A;  $p > 0,05$ , prueba Mann-Whitney  $U$ ).

A los 6 meses, no contábamos con datos suficientes para sacar conclusiones firmes. Se observó una tendencia a una mayor carga leucémica en el grupo de pacientes con peor respuesta (Fig. 3.8, C-D).

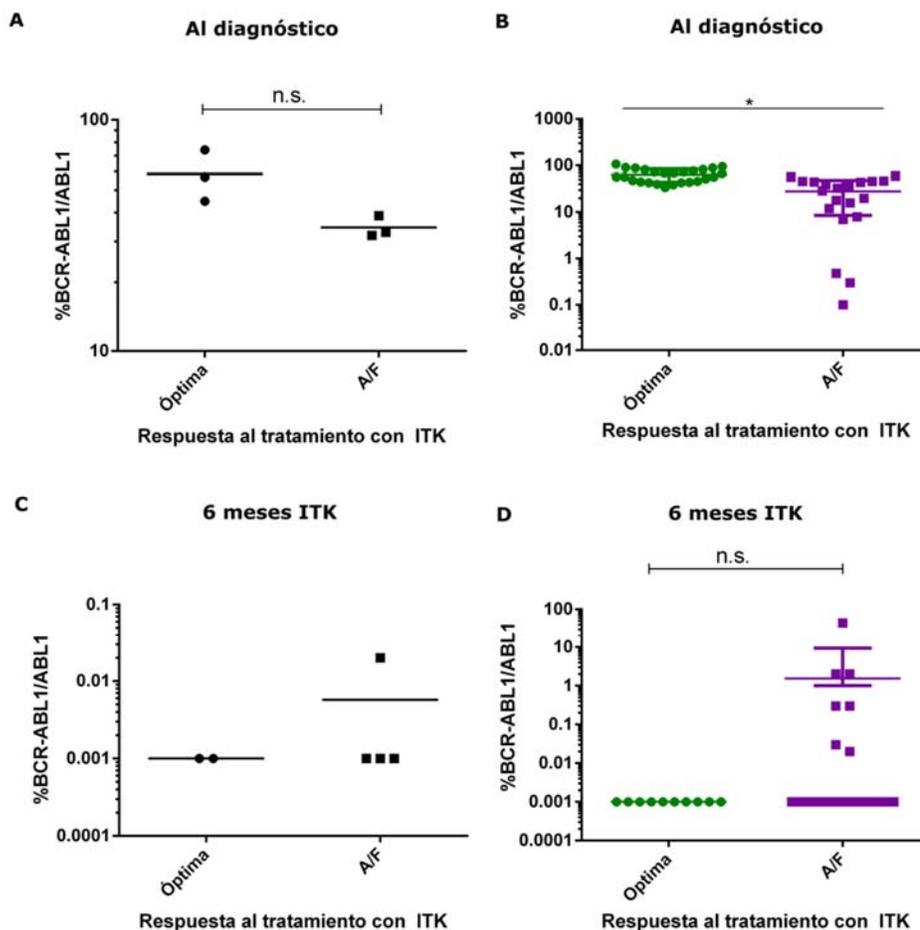


Fig. 3.8. Comparación de la carga leucémica en la fracción progenitora mediante ensayos CFU con muestras provenientes de pacientes al diagnóstico, previo tratamiento (A-B) o a los 6 meses de tratamiento con ITK 1ª línea (C-D). \*= $p < 0,05$ ; A/F = "advertencia/falla"

### III.5. Detección de LSC basada en la amplificación del punto de fusión *BCR-ABL1* a nivel de ADN genómico.

Considerando a las LSC como responsables de la persistencia de la enfermedad en pacientes bajo tratamiento prolongado con ITK, resulta de especial interés su detección en pacientes con RM profunda, buscando responder las siguientes preguntas:

-¿existen distintos niveles de LSC residuales en estos pacientes?

-Si es así, ¿correlacionan con la capacidad de discontinuar el tratamiento en forma segura?

En la mayoría de los pacientes con respuesta RM4.0-RM5.0 no pudimos detectar colonias positivas para *BCR-ABL1* en el ensayo de células primitivas (Tabla 3.3). Dado que la metodología empleada dependía de los niveles de expresión de *BCR-ABL1*, nos propusimos desarrollar una metodología que nos permita cuantificar las LSC a nivel de ADN genómico (ADNg), basada en la detección del punto de fusión de cada paciente.

Tabla 3.3. Porcentaje de pacientes con al menos una colonia positiva para el ARNm *BCR-ABL1*.

Respuesta molecular en SP	LTC-IC de 6 semanas ( <i>BCR-ABL1</i> +)
RM <sup>0</sup> (sin tratamiento)	100% (8/8)
RM <sup>1.0</sup> -RM <sup>3.0</sup>	84% (16/19)
RM <sup>4.0</sup> -RM <sup>5.0</sup>	20% (1/5)

#### III.5.1. Puesta a punto de la secuenciación del punto de ruptura *BCR-ABL1* a nivel genómico

Los sitios de ruptura genómica que llevan a la formación del gen quimera *BRC-ABL1* se pueden localizar a lo largo de una región de ~3kpb en *BCR* (intrón 13 o intrón 14), y una región de ~140kpb en *ABL1* (intrón 1) (Fig. 3.9). Por lo tanto, cada paciente posee, a nivel genómico, un punto de ruptura específico, el cual es necesario conocer para el diseño de *primers* para PCR.

Con el objetivo de secuenciar el punto de ruptura de cada paciente, se utilizó la metodología de secuenciación masiva dirigida, a partir de amplicones. Para la obtención de los amplicones se utilizó una estrategia descrita en bibliografía<sup>75</sup>, basada en una PCR *long-range*. Ésta utiliza un *primer forward* que hibrida en el exón 13 o 14 de *BCR* (según el rearreglo sea b2a2 o b3a2, respectivamente), y 19 *primers reverse* que se distribuyen a lo largo del intrón 1 de *ABL1*, cada aproximadamente 500pb (exceptuando un *primer*, que hibrida en el intrón 3 de *ABL1*), divididos en 4 reacciones *multiplex* (Fig. 3.9, Anexo II). El tamaño del producto de PCR es desconocido; en nuestro caso, las muestras que pudieron ser exitosamente secuenciadas dieron productos de PCR de longitud promedio 4078pb

(rango: 770-8000pb). La banda obtenida se purificó del gel y se utilizó para la construcción de la biblioteca para secuenciación masiva (ver Materiales y métodos).

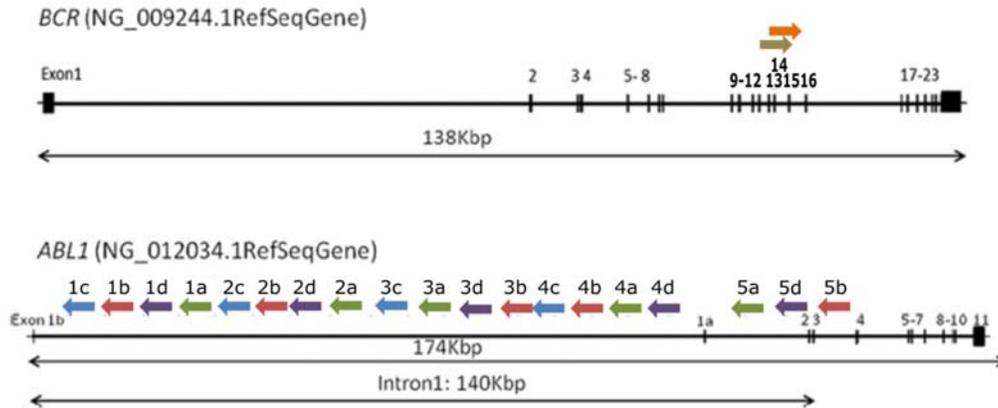


Fig. 3.9. Descripción de la estrategia de PCR long-range multiplex para la amplificación del punto de ruptura BCR-ABL1 a nivel genómico. Se indican las posiciones relativas de los primers reverse en ABL1 pertenecientes a cada set (A, B, C, D), y los primers forward en los exones 13 o 14 de BCR. Modificado de <sup>84</sup>.

Se utilizaron dos plataformas de secuenciación: Ion Torrent PGM (Thermo Fisher Scientific) y MiSeq (Illumina). Las mismas difieren en sus características, y en el protocolo de construcción de la biblioteca (Tabla 3.4):

Tabla 3.4. Características de las metodologías de secuenciación masiva empleadas para la detección del punto de ruptura BCR-ABL1 en ADN genómico. \*=tamaño medio de las lecturas obtenidas en la corrida; el tamaño teórico esperado era de 400pb.

	Ion Torrent PGM	MiSeq
<b>Química de detección</b>	Cambio de pH	Nucleótidos bloqueados 3' reversibles fluorescentes + <i>imaging</i>
<b>Amplificación</b>	PCR en emulsión	PCR puente
<b>Tamaño de las lecturas</b>	170pb*( <i>single-end</i> )	2 x 250pb ( <i>paired-end</i> )
<b>Errores sistemáticos</b>	Regiones homopoliméricas	Sub-representación de regiones ricas en G-C o A-T

La primera corrida se realizó en la plataforma Ion Torrent PGM. La mayoría de los algoritmos bioinformáticos disponibles para la detección de variantes estructurales utilizan, a partir de las lecturas provenientes de ambos extremos de una secuencia (diseño *paired-end*), la discordancia en distancia y/o orientación (Fig. 3.10). Sin embargo, la secuenciación en la plataforma Ion Torrent PGM es *single-end*. Por ello, se utilizó el algoritmo CREST (Clipping REveals STStructure)<sup>77</sup>, el cual permite la utilización de lecturas provenientes de secuenciaciones *single-end*, basándose en la premisa de que, si se cuenta con lecturas lo suficientemente largas ( $\geq 75$ pb), una lectura proveniente de una variante estructural tendrá una parte que alinea con una región del genoma, y otra parte de la lectura que alinea con otra región del genoma. Realizando una combinación de las porciones de lecturas no alineadas, se puede alinear nuevamente ese ensamble al genoma

de referencia, e identificar la región genómica translocada (Fig. 3.10). De las 9 muestras secuenciadas, 3 no pasaron el control de calidad, y se logró obtener la secuencia de 4 pacientes y 2 líneas celulares.

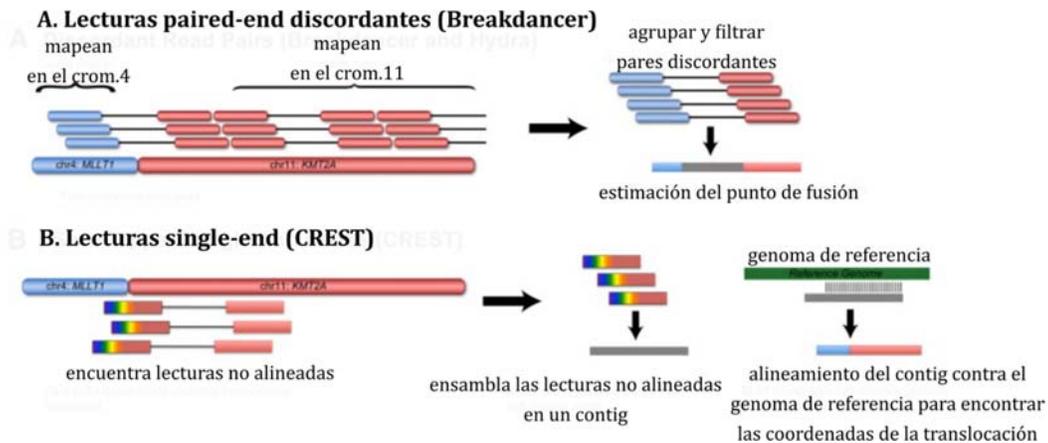


Fig. 3.10. Estrategias bioinformáticas para la identificación de translocaciones con datos provenientes de secuenciación masiva. A: método con el que se analizaron los datos provenientes de la secuenciación en MiSeq (Illumina); B: método con el que se analizaron los datos provenientes de Ion Torrent PGM. Modificado de <sup>85</sup>.

La segunda corrida se realizó en la plataforma MiSeq. En este caso fue posible probar una mayor variedad de algoritmos bioinformáticos, ya que la secuenciación fue *paired-end*. Se utilizó el algoritmo BreakDancer<sup>78</sup>, el cual se basa en la discordancia entre lecturas provenientes de la misma secuencia, como se explicó previamente (Fig. 3.10). En este caso se logró obtener el punto de fusión de 17/19 muestras de pacientes, y de 2 líneas celulares. Además, 8 de las 21 secuencias fueron obtenidas automáticamente.

Las secuencias obtenidas de las líneas celulares (K562 y MEG-01) coincidieron con las reportadas en bibliografía<sup>86</sup>. Cinco muestras de pacientes habían sido previamente caracterizadas por otro método en la Academia Nacional de Medicina (PCR inversa de larga distancia), y 4/5 coincidieron en la secuencia obtenida por ambos métodos. El caso discordante fue secuenciado por Ion Torrent y MiSeq; se trata de un caso atípico, ya que el análisis por Ion Torrent arrojó dos tipos de rearrreglos, diferentes del resultado obtenido por MiSeq. El paciente presentó resistencia a ITK de 1<sup>ra</sup> línea, y aún no se ha podido verificar la secuencia mediante *primers* específicos.

### III.5.2. Detección de la carga leucémica en la fracción primitiva a nivel de ADN genómico.

Se seleccionaron cuatro pacientes para el diseño de *primers* específicos, y la detección del clon leucémico en la fracción primitiva (Tabla 3.5, Fig. 3.11, Anexo II). Los sitios de ruptura del cromosoma se correspondieron con lo esperado según la isoforma del ARNm en los 4 casos. En el caso MP5920 se detectó una inversión de 75pb de *ABL1* en el sitio de ruptura, la cual también fue detectada por la metodología empleada en la Academia Nacional de Medicina. En el caso N22 se detectó el punto de ruptura en el exón 14 (Fig. 3.11). Dado lo atípico de este caso, la secuencia fue verificada por la metodología

de Sanger. Según lo reportado en bibliografía, en estos casos el exón resulta eliminado del transcripto maduro por el proceso de corte y empalme<sup>8</sup>.

*Tabla 3.5. Características de los pacientes para los cuales se realizó el ensayo de detección de células BCR-ABL1+ a nivel de ADNg en la fracción de células primitivas.*

<b>Código</b>	<b>Edad al Dx</b>	<b>Sokal</b>	<b>Isoforma ARNm</b>	<b>Set de primers PCR long range</b>	<b>Plataforma de secuenciación</b>
<b>N22</b>	66	Intermedio	b2a2	BCRB2F - Set B	Ion Torrent
<b>N38</b>	31	Alto	b3a2	BCRB3F - Set B	MiSeq
<b>N35</b>	24	Intermedio	b3a2	BCRB3F - Set C	MiSeq
<b>MP5920</b>	33	Bajo	b2a2	BCRB2F - Set B	Ion Torrent y MiSeq



-extracción de ADNg por columna a partir de pooles de colonias, seguido de qPCR (un método análogo al utilizado para la detección de ARNm en ensayos LTC-IC);

-lisados obtenidos a partir de colonias individuales, sin un paso de extracción de ADNg, sometidos a qPCR;

-lisados obtenidos a partir de colonias individuales, sin un paso de extracción de ADNg, sometidos a PCR *nested/semi-nested*.

La calidad del material fue evaluada por amplificación de un gen independiente (B-RAF). Se utilizó como control positivo de amplificación de *BCR-ABL1*, ADNg extraído al diagnóstico, y como control negativo, una región de la metilcelulosa circundante a las colonias.

En primer lugar, dado que la motivación para realizar este ensayo radicaba en la baja capacidad de detección a nivel de ARNm, se comparó el porcentaje de pooles de colonias *BCR-ABL1* positivas mediante la detección de ARNm vs. ADNg. Los resultados obtenidos sugirieron una mejora en la sensibilidad mediante la detección a nivel de ADNg por PCR punto final, ya que el 100% de los pooles de colonias analizados resultaron positivos vs. un 60-67% por ARNm, tanto en ensayos de corto como largo término (Tabla 3.6). A su vez, los pooles que resultaron positivos para el ARNm de *BCR-ABL1* presentaron valores cercanos al límite de detección, sugiriendo que en pacientes con RM más profunda la cuantificación del ARNm no sería posible (como se vio en la sección anterior).

Tabla 3.6. Comparación de la capacidad de detección de pooles de colonias positivas para *BCR-ABL1* a nivel de ADNg (PCR punto final) y ARNm. Entre paréntesis se indica el valor promedio del Ct para los pooles positivos.

<b>Ensayo LTC-IC N22 9m</b>		
	ADNg	ARNm
<b>No. pooles positivos <i>BCR-ABL1</i></b>	3/3	2/3 (Ct =38,8)
<b>Ensayo CFU, N22 12m</b>		
	ADNg	ARNm
<b>No. pooles positivos <i>BCR-ABL1</i></b>	5/5	3/5 (Ct = 40,1)

Con el objetivo de simplificar el método de *screening*, se puso a punto la PCR directa a partir de colonias individuales. El método consiste en lisar las colonias en un buffer con detergente y proteinasa K, mediante calentamiento (ver Materiales y métodos). Una parte del lisado (1/10 del volumen) fue utilizada como molde de la PCR. Se comparó la capacidad de detección de la qPCR vs. PCR *nested/semi-nested*. En estas condiciones, la amplificación por PCR *nested* mejoró la sensibilidad del ensayo en comparación con qPCR (Tabla 3.7).

Tabla 3.7. Comparación de la capacidad de detección de colonias positivas para BCR-ABL1 a nivel de ADNg entre qPCR y PCR nested.

Ensayo CFU, N22 6m		
	qPCR	PCR nested
No. lisados positivos BCR-ABL1	4/9	9/9
No. lisados positivos B-RAF	9/9	
Ensayo LTC-IC, N22 18m		
	qPCR	PCR nested
No. lisados positivos BCR-ABL1	0/20	5/20
No. lisados positivos B-RAF	20/20	

Se observó que para la detección de *B-RAF* fue suficiente con la metodología de qPCR o PCR tradicional (1 ronda), aún en lisados individuales. Descartamos que esta diferencia se deba a la eficiencia de amplificación, ya que la misma, estimada por qPCR, fue similar para ambos genes (Fig. 3.12). Sin embargo, la sensibilidad fue menor para *BCR-ABL1*, ya que los puntos más diluidos resultaron indetectables, mientras que en *B-RAF*, una de las réplicas levantó en Ct=36,1 mientras que la otra resultó indetectable. En estas condiciones, el rango detectable fue de 20pg-20ng para *BCR-ABL1* y de 2pg-20ng para *B-RAF*.

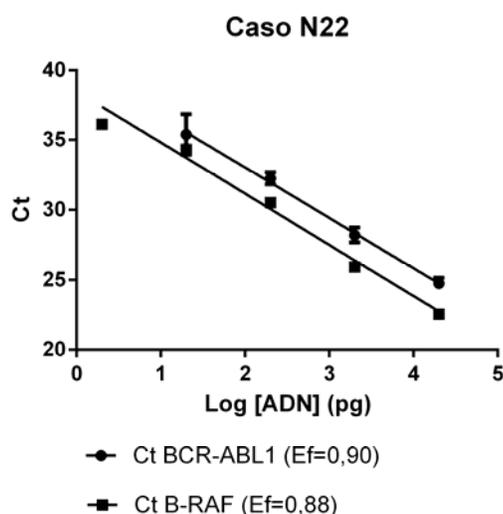


Fig. 3.12. Curvas de eficiencia para la amplificación por qPCR de BCR-ABL1 y B-RAF a nivel de ADNg genómico con una muestra al diagnóstico del caso N22.

Cabe notar que esta curva se realizó con una muestra de MO al diagnóstico, en la cual el porcentaje de células *BCR-ABL1+* no tiene que ser necesariamente el 100%, mientras que sí deberían ser todas portadoras de dos alelos de *B-RAF*. Consideramos sin embargo que éste no es el factor más importante para explicar la diferencia en la sensibilidad, ya que los valores de Ct obtenidos por qPCR a partir de colonias (en las cuales se espera que el 100% de las células tengan el mismo genotipo) dieron resultados en el mismo sentido: los valores de Ct para *BCR-ABL1* fueron mayores que para *B-RAF* ( $\Delta Ct$  promedio = 4,7).

La metodología de PCR directa *nested* o *semi-nested* sobre lisados de colonias individuales fue aplicada al *screening* de pacientes en RM 4.0, 4.5 y 5.0, es decir, con respuestas más profundas. La sensibilidad del ensayo paciente-específico fue verificada mediante curvas de masa, utilizando una muestra de cada paciente al diagnóstico, siendo la mínima cantidad detectada por PCR *nested/semi-nested* de 200pg (N38) y 20pg (N35) (Fig. 3.13). En el caso de MP5920, el ensayo no fue óptimo ya que se detectaron productos inespecíficos al analizar lisados individuales, y baja sensibilidad (2ng). Dado que se trata de una muestra de ingreso reciente, próximamente diseñaremos nuevos *primers* para poder mejorar la especificidad y sensibilidad.

Se observó una disminución en la frecuencia de colonias leucémicas en los pacientes con menor nivel de enfermedad residual (Tabla 3.8). Dado que el límite de detección del método está determinado por el número de colonias analizadas, concluimos que en pacientes con RM profunda (RM 4.5-RM5.0) es necesario aumentar el número total de colonias analizadas.

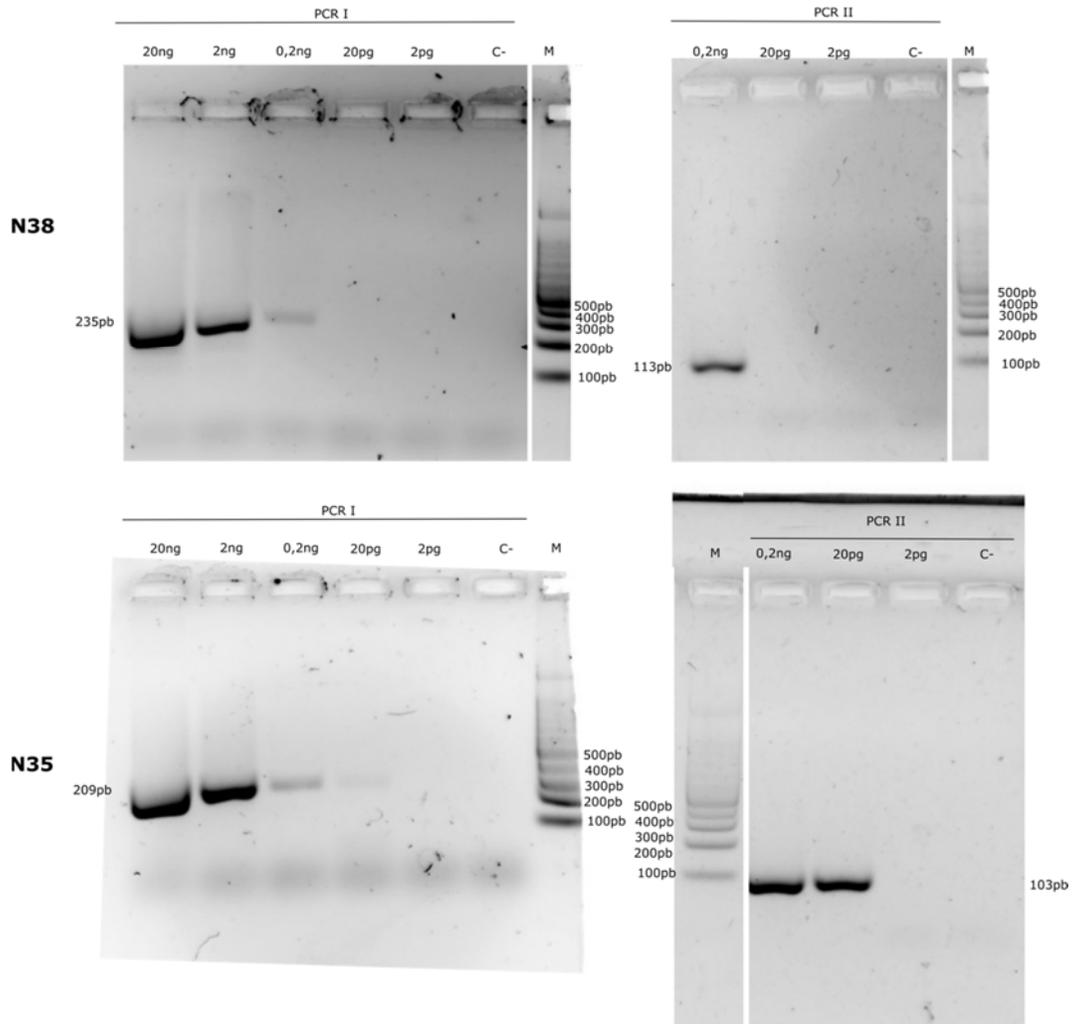


Fig. 3.13. Evaluación de la sensibilidad del método de PCR *nested/semi-nested* para la detección del punto de fusión de N38 y N35. Se utilizó como molde una muestra de ADN genómico extraído al diagnóstico, previo tratamiento con *ITK*.

Tabla 3.8. Cuantificación de la carga leucémica en la fracción primitiva a nivel de ADNg.

	Tiempo de tratamiento con ITK (años)	ITK	Respuesta molecular	Tiempo de sostenimiento de última RM	Lisados BCR-ABL+ (%)
<b>N22</b>	1,5	Imatinib	RM 3.0	≥6 meses	5/20 (25%)
<b>N38</b>	1,6	Nilotinib	RM 4.0	≥7 meses	2/20 (10%)
<b>N35</b>	1,7	Dasatinib	RM 4.5	8 meses	1/19 (5%) (pendiente de secuenciación)
<b>MP5920</b>	3,7	Imatinib	RM 5.0 indetectable	≥2 años (RM4.5-RM5.0)	0/20 (0%)

### III.6. Conclusiones parciales y comentarios

A continuación, se discuten las principales conclusiones extraídas a partir de los resultados de esta primera parte:

1) *Las células primitivas provenientes de pacientes al diagnóstico presentan menor capacidad de mantenimiento a largo plazo en cultivos in vitro, sugiriendo una disminución en su capacidad de autorrenovación. No se detectaron diferencias en la capacidad proliferativa de células progenitoras.*

Los ensayos clonogénicos de corto término (CFU) son utilizados ampliamente no sólo en el campo de la investigación biomédica sino también para evaluar la capacidad proliferativa de células progenitoras criopreservadas, como es el caso de los bancos de células de cordón umbilical. La capacidad de formar colonias en este ensayo es uno de los parámetros a tener en cuenta al momento de evaluar si una muestra es apta o no como fuente para un trasplante de células precursoras hematopoyéticas. En nuestros ensayos CFU no detectamos diferencias en la capacidad proliferativa de la fracción de células progenitoras entre pacientes al diagnóstico y bajo tratamiento con ITK. Este resultado nos llamó la atención ya que los pacientes al diagnóstico se caracterizan por tener una alta proliferación de células del linaje mieloide. Una posible explicación es que el ensayo, al restringir la superficie de crecimiento (los ensayos fueron realizados en placas p35), imponga una cota superior en el número total de colonias observadas. Otra posibilidad es que las condiciones de crecimiento requeridas por las células progenitoras leucémicas no sean las mismas que las requeridas por células progenitoras normales, para las cuales se ha desarrollado este ensayo.

En cambio, en ensayos de largo término (LTC-IC) se observó una menor capacidad proliferativa de las células provenientes de pacientes al diagnóstico y en RM 1.0 (Fig. 3.1). Este ensayo se basa en la premisa de que las células primitivas pueden sobrevivir múltiples divisiones celulares y aún no agotarse, mientras que las células más diferenciadas agotarían su ciclo de vida a lo largo de las 6 semanas de cultivo. Dado que en pacientes al diagnóstico y RM 1.0 la carga leucémica es mayor que en aquellos con mayor tiempo de tratamiento (Fig. 3.6), este resultado sugiere que la capacidad de autorrenovación estaría disminuida en la fracción de células primitivas leucémicas. Esta observación está de acuerdo con lo reportado por Udomsakdi *et al*, trabajo en el que se

realizaron ensayos LTC-IC con muestras de LMC al diagnóstico y de donantes sanos, detectándose una menor capacidad proliferativa en las primeras<sup>87</sup>. En estudios más recientes, donde se caracterizó el patrón de expresión génica de células primitivas de pacientes con LMC y donantes sanos<sup>88</sup>, se observó una menor distancia fenotípica entre las células primitivas leucémicas y sus contrapartes más diferenciadas, que entre las células primitivas normales y sus contrapartes más diferenciadas. Es decir, las células primitivas leucémicas tendrían menor "stemness" o características de células madre. Este fenómeno es de interés para la evaluación de lo que ocurre *in vivo* en los pacientes: si las células primitivas leucémicas requieren menos ciclos de replicación para agotarse en comparación con las células primitivas normales, implicaría que la eliminación completa del clon leucémico es teóricamente factible. Esta hipótesis concuerda con la observación extraída de ensayos de discontinuación de ITK, en los que se detectó a la variable "tiempo de duración de la RM profunda" como factor pronóstico de la posibilidad de mantener el TFR<sup>89</sup>.

*2) El compartimiento de células primitivas y progenitoras responde al tratamiento con ITK in vivo, siendo que la carga leucémica, estimada a través de ensayos funcionales in vitro, disminuye a lo largo del tiempo.*

La disminución de la fracción de células primitivas a lo largo del tratamiento contrasta con los resultados provenientes de ensayos *in vitro*, en los que se observó que células de la fracción primitiva de pacientes al diagnóstico, a pesar de estar tratadas con ITK, podían crecer en presencia y ausencia de factores de crecimiento<sup>51</sup>. Ésto podría deberse a componentes extrínsecos no considerados en los ensayos *in vitro*, y/o a que los tiempos de exposición a ITK en los ensayos *in vitro* son muy breves (en general 24-72hs).

*3) La dinámica de la fracción de células primitivas una vez iniciado el tratamiento correlaciona con la respuesta observada en células más diferenciadas de SP y MO, siendo que aquellos casos de menor respuesta al tratamiento con ITK también presentaron una menor disminución relativa de los niveles de carga leucémica en la fracción primitiva a los 6 meses de tratamiento.*

El ensayo utilizado para la detección de células primitivas tiene como principal limitación que requiere mucho tiempo para evaluar el resultado (>2 meses). Ésto implica que su aplicación como complemento del monitoreo de los pacientes sería complejo, especialmente dada la observación de que los signos de alarma aparecen durante los primeros meses de tratamiento. Sin embargo, los resultados contribuyen a comprender la biología de la enfermedad, ya que sugieren que lo que ocurre en el compartimiento de células primitivas refleja (u origina) lo que ocurre en la fracción de células más diferenciadas.

*4) La carga leucémica en la fracción de células progenitoras al diagnóstico fue mayor en pacientes con mejor respuesta al tratamiento con ITK.*

Dado que son pocos casos los que pudimos evaluar, hacen falta más pacientes para poder sacar conclusiones generales sobre los patrones observados. Cabe notar que, aún siendo una cohorte pequeña, la proporción de casos categorizados como advertencia/falla fue de 7/16 (44%) (Tabla 3.2). Los resultados obtenidos podrían interpretarse preliminarmente tal que, al diagnóstico, aquellos pacientes con mayores niveles de carga leucémica tendrían

mayor dependencia con la actividad quinasa de BCR-ABL1, y por lo tanto serían más susceptibles al tratamiento con ITK. Por el contrario, en aquellos pacientes con menores niveles de carga leucémica, las células leucémicas serían menos dependientes de la actividad quinasa de BCR-ABL1, y por lo tanto menos susceptibles al tratamiento con ITK. En un estudio reciente se observó que, en una cohorte de 29 pacientes, los 3 casos que presentaron menores niveles de transcriptos *BCR-ABL1* en SP al diagnóstico, luego presentaron fracaso del tratamiento (2/3) o signos de advertencia (1/3). En dicho trabajo también se realizó FISH en las muestras al diagnóstico, y 2 de los 3 casos presentaron casi un 100% de células Ph<sup>+</sup> (1 de los casos fue "no analizable"), indicando que los bajos niveles de ARNm *BCR-ABL1* no eran debidos a una menor frecuencia del clon leucémico<sup>90</sup>.

*5) La sensibilidad de la metodología empleada para la detección de células primitivas leucémicas a nivel de ARNm no fue suficiente para aquellos pacientes bajo tratamiento prolongado con ITK; requiriendo del desarrollo de nuevos métodos de screening. Fue posible secuenciar el punto de fusión BCR-ABL1 a nivel de ADNg de forma exitosa en 20 muestras de pacientes, permitiendo el diseño de primers paciente-específicos.*

El objetivo de esta parte del proyecto fue desarrollar una metodología para la detección de LSC en pacientes bajo tratamiento con ITK, que no dependa de los niveles de expresión de *BCR-ABL1*, ya que éstos pueden verse disminuidos por la presión ejercida por el tratamiento, seleccionando clones menos sensibles a los ITK<sup>91</sup>. Por otro lado, el ARN es más susceptible de ser degradado que el ADN y requiere de cuidados especiales para su almacenamiento. La secuenciación del punto de ruptura específico de cada paciente nos permitió diseñar *primers* para evaluar las colonias obtenidas en ensayos LTC-IC mediante una PCR directa. La ventaja de la PCR directa radica en un ahorro sustancial de costos, ya que evita la necesidad de usar kits comerciales para extracción de ADNg a partir de muestras poco abundantes (como es el caso de colonias individuales o pooles de colonias). Por otro lado, la amplificación que otorgó la PCR *nested* o *semi-nested* mejoró la capacidad de detección del método en comparación con la metodología de qPCR. Desconocemos los motivos de la baja sensibilidad de la qPCR para *BCR-ABL1* a nivel de ADNg, ya que la eficiencia del ensayo fue comparable a la de *B-RAF*.

La principal limitación de la metodología es la necesidad de optimizar el ensayo de PCR *nested* para cada paciente. A su vez, la región genómica correspondiente al punto de fusión está enriquecida en regiones repetidas, secuencias palindrómicas, regiones de micro-homología, y bajo contenido de GC en *ABL1*, que dificultan el diseño de *primers* y la optimización de una PCR de alta sensibilidad y especificidad<sup>8</sup>.

*6) La detección de células de la fracción primitiva a nivel de ADNg mejoró la capacidad de detección del método en comparación con la detección del ARNm, permitiendo el screening de pacientes en RM3.0, RM4.0 y RM4.5.*

Consideramos a los resultados obtenidos como una prueba de concepto de que la detección de LSC mediante ensayos funcionales en pacientes con RM profunda, a nivel de ADNg, es factible. Con la optimización adecuada para cada paciente y un aumento en el número de colonias totales evaluadas, creemos que sería posible aplicar esta metodología a pacientes candidatos a discontinuación de ITK. Con el objetivo de evitar un aumento sustancial de costos y trabajo, se podría realizar la recolección de más colonias

individuales, pero que sean inicialmente evaluadas en pooles. Conservando el material original perteneciente a dichos lisados, el objetivo sería identificar pooles con al menos una colonia positiva, que luego podrían ser reevaluados, repitiendo la PCR con las colonias individuales por separado. Mediante este método se podrían mantener las condiciones que ya fueron optimizadas.

## Capítulo IV. Caracterización y validación del miRNoma de LSC y HSC de pacientes con LMC al diagnóstico.

### IV. 1. Introducción

La caracterización a nivel molecular de la fracción de células primitivas presentes en pacientes con LMC ha cobrado gran interés a partir de las evidencias que muestran su resistencia a los efectos pro-apoptóticos del tratamiento con ITKs<sup>50,53,54</sup>, e incluso la selección de una subpoblación quiescente, de baja expresión de *BCR-ABL1*, durante el tratamiento prolongado<sup>55,91</sup>. Con el objetivo de encontrar marcadores específicos de esta población, se han logrado grandes avances mediante el estudio de la expresión diferencial de genes por *microarrays* y RT-qPCR, los cuales han descrito:

- expresión aberrante de un conjunto de marcadores de superficie: CD26, CD25, IL-1RAP, CD9, CD33, CD36, CD44, CD52, CD90, CD105, CD114, CD117, CD123, CD164, CD184, ROBO4, NPDC1<sup>92-94</sup>;

- sobrexpresión de genes involucrados en ciclo celular y proliferación, procesamiento de ARNm, traducción, reparación de daño al ADN, metabolismo oxidativo, procesamiento de proteínas, respuesta inmune y procesos metabólicos<sup>93</sup>;

- subexpresión de genes asociados a matriz extracelular, superficie celular, programas de diferenciación y desarrollo, respuesta a estímulos, vías de señalización de TGF- $\beta$  (*Transforming growth factor beta*) y BMP (*Bone morphogenetic proteins*)<sup>93</sup>;

- un perfil transcripcional correspondiente a células más diferenciadas que sus contrapartes normales, con capacidad defectuosa de adhesión y migración *in vitro*, subexpresión de factores de transcripción y moléculas involucradas en el mantenimiento de la multipotencia, sugiriendo una pérdida de quiescencia<sup>88</sup>;

- dependencia alterada de la señalización vía MYC/TP53 a través de la caracterización del proteoma<sup>95</sup>.

Este tipo de abordajes tienen como principal desafío la descripción de marcadores o procesos biológicos que no sean compartidos por sus contrapartes normales, buscando minimizar los potenciales efectos tóxicos de nuevas combinaciones terapéuticas. A su vez, el compartimiento definido como LSC es heterogéneo y de baja frecuencia<sup>96,97</sup>.

La fracción de células primitivas, conteniendo a las HSC y LSC, corresponde a la población Lin-CD34<sup>+</sup>CD38<sup>-</sup>. Entre los marcadores de superficie descritos como diferencialmente expresados en LSC de LMC (Fig. 4.1), CD26 resulta de especial interés por su ausencia en el compartimiento normal y en otras patologías del sistema hematopoyético. Se trata de una peptidasa de membrana (DPPIV, dipeptidilpeptidasa-IV) capaz degradar al factor quimiotáctico SDF-1 involucrado en el reclutamiento de células CXCR4<sup>+</sup> a MO<sup>94</sup>.

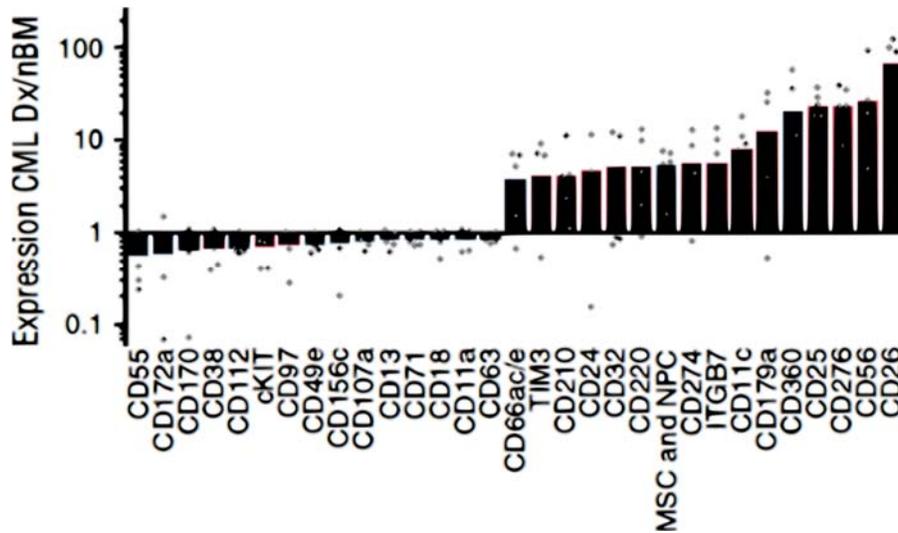


Fig. 4.1. Expresión relativa de marcadores de superficie medidos en la población de LSC de MO de pacientes con LMC al diagnóstico y en MO de dadores sanos. Extraído de <sup>97</sup>.

Con el objetivo de ampliar el conocimiento sobre los procesos biológicos que ocurren diferencialmente en la población de LSC, hemos realizado una caracterización del miRNoma por secuenciación masiva. Para ello se compararon LSC de pacientes con LMC al diagnóstico con su contraparte normal presente en pacientes y en dadores sanos (DS).

#### IV. 2. Puesta a punto del aislamiento de la fracción de LSC y HSC de pacientes con LMC al diagnóstico

Con el objetivo de estudiar el miRNoma de las HSCs y LSCs, fue necesario poner a punto su aislamiento a partir de muestras de pacientes con LMC. Para ello se utilizó separación celular basada en citometría de flujo (FACS *sorting*), combinando la expresión de los marcadores reportados en bibliografía. Mientras que en el trabajo de Janssen *et al* utilizaron los niveles de intensidad de fluorescencia media de los parámetros CD45, CD34, FSC y SSC para distinguir a ambas poblaciones<sup>98</sup>, en el trabajo de Herrmann *et al* describieron al marcador CD26 como exclusivo de la fracción de LSC, correlacionando con la presencia del cromosoma Ph<sup>+</sup> y con los niveles de ARNm de *BCR-ABL1* al diagnóstico<sup>94</sup>. En nuestras manos, cualquiera de las dos estrategias por separado no resultó en una clara distinción entre HSC y LSC en todos los casos evaluados, resultando muy subjetiva la definición de las poblaciones. Sin embargo, con la combinación de ambas estrategias obtuvimos una separación eficiente, teniendo en cuenta las siguientes consideraciones:

- La línea de corte para el marcador CD38 debe ser rigurosamente establecida utilizando un control de isotipo, sobre la población de células CD45<sup>low</sup>. Esto evita la contaminación con células progenitoras CD38<sup>dim</sup>, las cuales expresan CD26 de forma indistinta con respecto a la presencia del cromosoma Ph<sup>+</sup>.
- No todos los pacientes presentan ambas poblaciones al diagnóstico. Se observaron 3 patrones (Fig. 4.2), que concuerdan parcialmente con los descriptos en bibliografía<sup>99</sup>. En aquellos pacientes con perfil 3, se observó la expresión de *BCR-ABL1* independientemente del nivel de expresión de CD26.

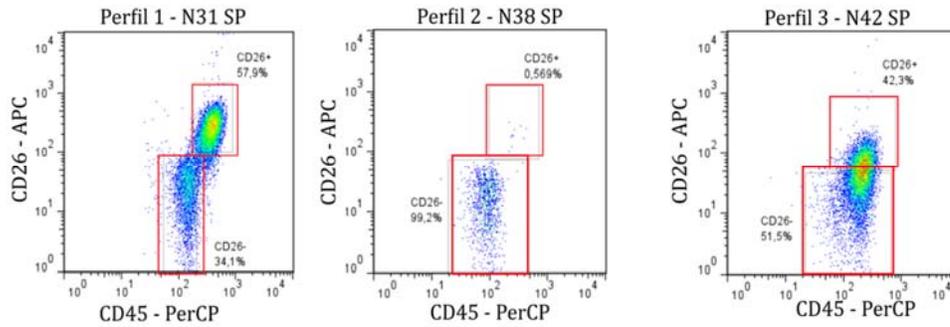


Fig. 4.2. Se muestran distribuciones representativas de los 3 perfiles observados en los pacientes al diagnóstico por citometría de flujo, con respecto a la proporción de las poblaciones CD26<sup>-</sup> y CD26<sup>+</sup>.

-El parámetro *window extension*, presente en el citómetro de flujo, debe ser fijado a 0. Esto evita que los abortos electrónicos contaminen las poblaciones aisladas.

El esquema de selección final se muestra en la figura 4.3. Las poblaciones fueron definidas como:

-**HSC**: CD45<sup>low</sup>, CD34<sup>+</sup>, CD38<sup>-</sup>, CD26<sup>-</sup>. Dentro de las CD34<sup>+</sup>CD38<sup>-</sup>, se comportan como CD45<sup>low</sup>, CD34<sup>low</sup>, FSC<sup>low</sup>, SSC<sup>low</sup>

-**LSC**: CD45<sup>low</sup>, CD34<sup>+</sup>, CD38<sup>-</sup>, CD26<sup>+</sup>. Dentro de las CD34<sup>+</sup>CD38<sup>-</sup>, se comportan como CD45<sup>high</sup>, CD34<sup>high</sup>, FSC<sup>high</sup>, SSC<sup>high</sup>

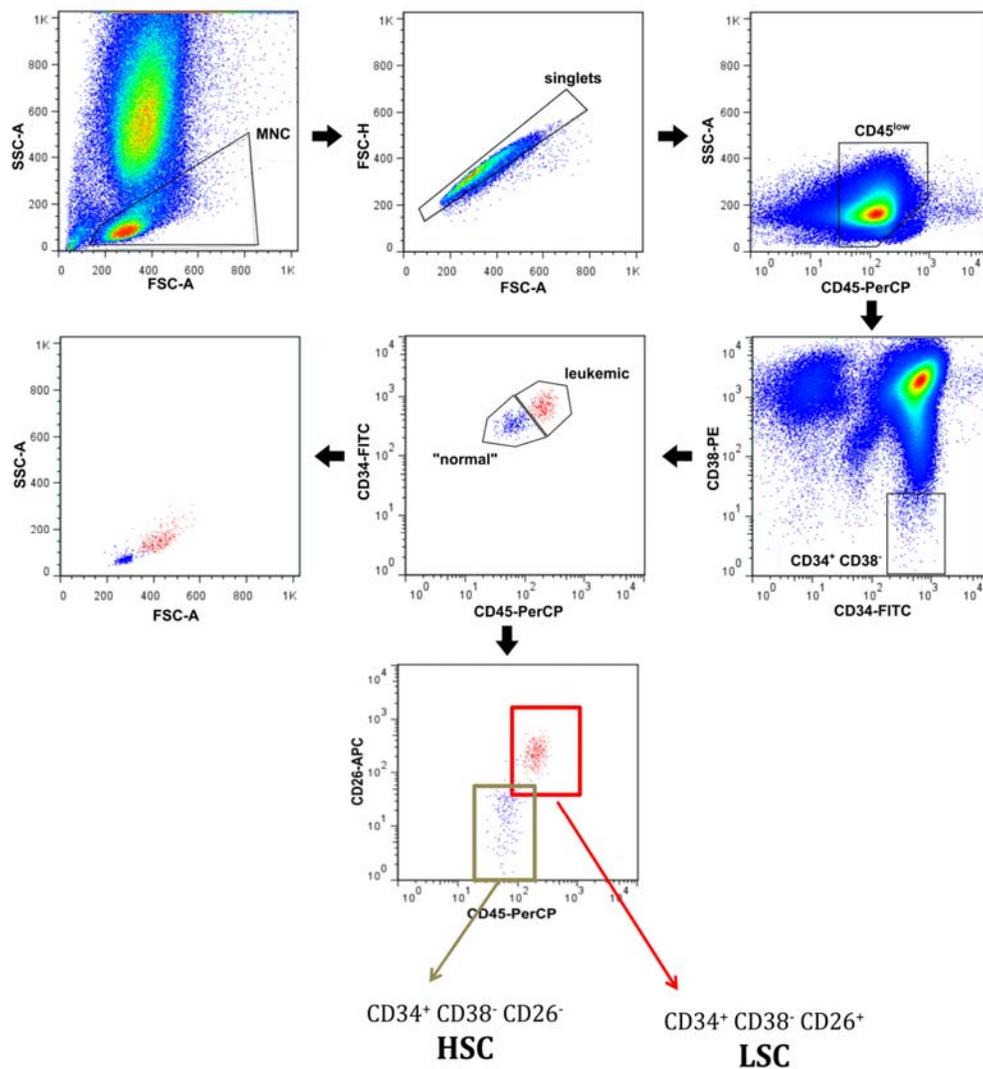


Fig. 4.3. Esquema de selección de poblaciones para el aislamiento de HSC y LSC a partir de muestras de SP o MO de pacientes con LMC al diagnóstico.

La pureza de las fracciones y la capacidad progenitora de las poblaciones aisladas fueron evaluadas mediante ensayos para progenitores (CFU), iniciados con células que fueron directamente recolectadas en el medio semisólido. A los 14-18 días, las colonias fueron extraídas (en pools de 3-6 colonias) para la determinación de los niveles del ARNm *BCR-ABL1* por RT-qPCR. Se obtuvieron pools 100% *BCR-ABL1*<sup>+</sup> en la fracción CD26<sup>+</sup> de pacientes con perfil 1 y 3, y 100% *BCR-ABL1*<sup>-</sup> en la fracción CD26<sup>-</sup> de pacientes con perfil 1 y 2. En el caso del perfil 3, las poblaciones CD26<sup>-</sup> y CD26<sup>+</sup> no se distinguían claramente, y ambas expresaban *BCR-ABL1* (Fig. 4.4). En estos pacientes sólo se utilizó la fracción CD26<sup>+</sup> en los ensayos subsiguientes.

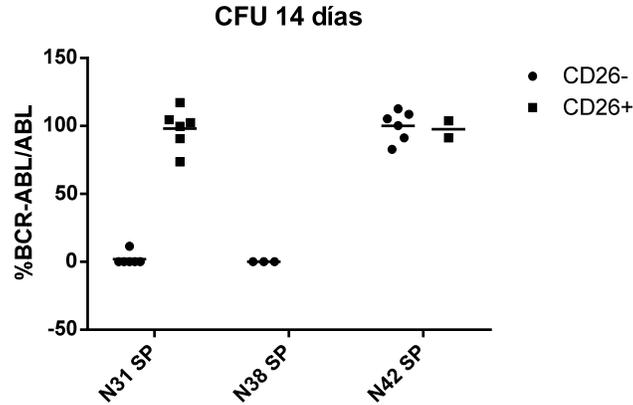


Fig. 4.4. Pureza de las fracciones obtenidas por citometría de flujo, medida retrospectivamente mediante cuantificación del ARNm BCR-ABL1 en ensayos CFU post-sorting.

La fracción de células primitivas se encuentra en el ápice del linaje hematopoyético, por lo tanto, su frecuencia es muy baja (mediana: 0,08% en MNC de SP de pacientes LMC al diagnóstico, n=9). Dado el bajo rendimiento por muestra (mediana: 479, rango: 22-9899 eventos), fue necesario optimizar la recolección, extracción y concentración de ARN conteniendo la fracción de ARN pequeños (<200nt). Se compararon 3 métodos de extracción de ARN (miRVana, RNAaqueous kit, RNAaqueous kit con modificación para conservación de la fracción de ARNs pequeños), y la preservación de la fracción de ARNs pequeños se evaluó mediante bioanalyzer. Se seleccionó el kit RNAaqueous con el protocolo modificado para la conservación de ARNs pequeños dado que el ARN obtenido fue de buena calidad, conservando la fracción de ARNs pequeños, y permitiendo eluir en un volumen menor que el método de miRVana (18µL vs 100µL) (Fig. 4.5).

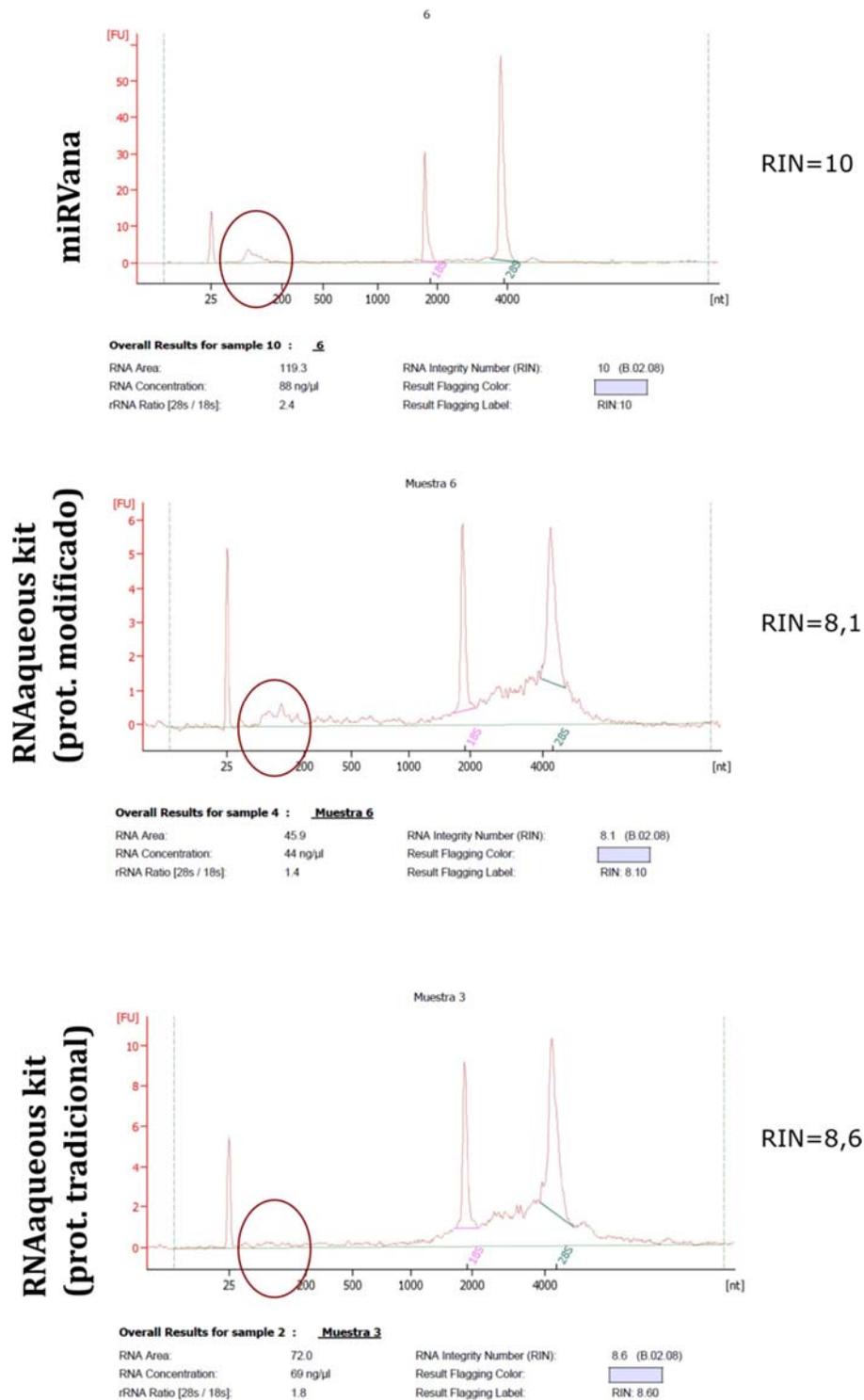


Fig. 4.5. Cromatograma de la corrida en Bioanalyzer (chip para ARN total Nano kit) de ARN total mediante el protocolo de miRVana, RNAaqueous-micro kit tradicional o modificado para la conservación de ARNs pequeños. Se indica con un círculo rojo la fracción de ARNs pequeños.

Se compararon las plataformas de secuenciación masiva disponibles (Ion Torrent e Illumina Hi-seq); seleccionamos la plataforma de Illumina dado que requería una menor

masa de ARN inicial (>100ng vs. >500ng), y porque la plataforma de Ion torrent presenta mayor tasa de errores sistemáticos (asociados a regiones con homopolímeros)<sup>100,101</sup>.

La cuantificación del ARN obtenido en las muestras individuales no fue posible ni por espectrofotometría (Nanodrop) ni por fluorescencia (Qubit), dado el bajo rendimiento individual por muestra. En consecuencia, se decidió realizar pools de muestras para poder reunir la masa de ARN necesaria para la construcción de bibliotecas para secuenciación masiva.

Por otro lado, el protocolo requería no sólo una masa mínima, sino una concentración mínima de ARN (20ng/µL). Se optimizó entonces un protocolo para la concentración de ARN en los pools de muestras. Para ello, una misma muestra fue dividida en dos, y sometida a centrifugación con vacío o a concentración por columna mediante un método comercial (ver Materiales y métodos). La conservación de la fracción de ARNs pequeños y la calidad del material fueron evaluados mediante bioanalyzer. La centrifugación con vacío fue superior a la utilización de columnas comerciales diseñadas especialmente para la concentración de ARN (Fig. 4.6), en lo que respecta a los parámetros evaluados.

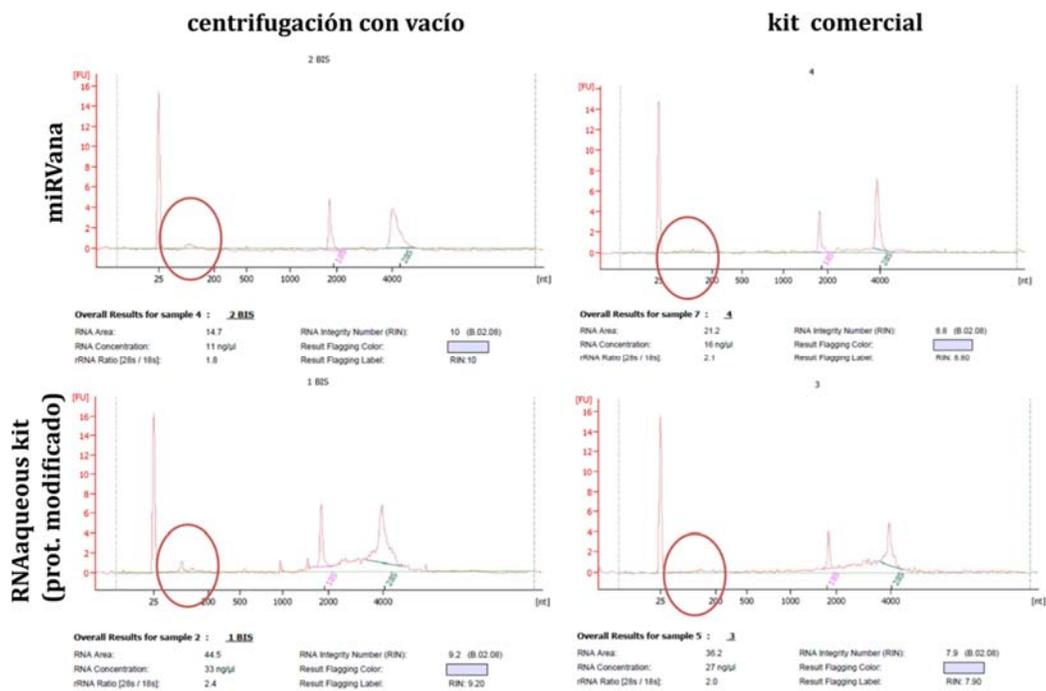


Fig. 4.6. Cromatograma de la corrida en Bioanalyzer (chip para ARN total Nano kit) de ARN total extraído a partir de células de MO de un paciente con LMC al diagnóstico (fracción CD34<sup>+</sup>), concentrado por centrifugación con vacío o por columna, para dos métodos de extracción de ARN. El círculo rojo indica la fracción de ARNs pequeños

Con respecto a los DS, se observó que el rendimiento de la fracción de células CD38<sup>+</sup> fue menor que para los pacientes con LMC (0,001% de MNC de SP, n=1). Por lo tanto, se tuvieron que realizar algunas modificaciones para llegar a reunir el número de eventos necesarios:

- se incluyó a la población CD38<sup>dim</sup> en la fracción de células primitivas
- se combinaron muestras provenientes de distintos DS.

Se aisló también la fracción correspondiente a las células progenitoras CD34<sup>+</sup>CD38<sup>+</sup> de DS como control positivo de una fracción abundante, sometida al mismo proceso de aislamiento por *sorting*, extracción de ARN y concentración por centrifugación con vacío

(Fig. 4.7). En este caso también se realizaron pools de dadores para mantener la variabilidad originada por la mezcla de muestras provenientes de distintas personas.

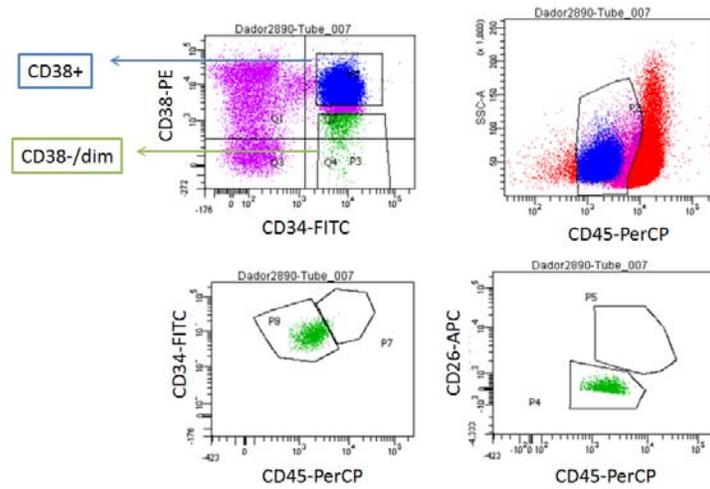


Fig. 4.7. Esquema de selección de poblaciones para aislamiento por sorting a partir de muestras de SP de DS.

### IV.3 Construcción de la biblioteca de ARNs pequeños para secuenciación masiva

Las muestras utilizadas para la construcción de la biblioteca para secuenciación masiva se describen en la Tabla 4.1.

Tabla 4.1. Características de las muestras utilizadas para la construcción de bibliotecas para secuenciación de ARNs pequeños (kit TruSeq, Illumina).

<b>Muestras LMC al Dx</b>						
<b>Código</b>	<b>Tipo de muestra</b>	<b>Sexo</b>	<b>Edad</b>	<b>Nro. Eventos HSC</b>	<b>Nro. Eventos LSC</b>	<b>Perfil</b>
<b>Pool 1</b>						
N32	SP	M	48	374	484	1
N32	MO	M	48	145	154	1
N31	SP	M	74	5562	9899	1
N31	MO	M	74	320	616	1
N22	SP	M	66	475	128	1
N23	MO	F	41	141	22	1
<b>Pool 2</b>						
N26	MO	M	24	6896	1857	1
N33	SP	M	54	1859	0	2
IM	MO	M	22	descartado	3046	3

<b>Dadores sanos</b>					
<b>Código</b>	<b>Tipo de muestra</b>	<b>Sexo</b>	<b>Edad</b>	<b>No. Evt CD38<sup>-</sup>/dim (HSC)</b>	<b>No. Evt CD38<sup>+</sup> (progenitores)</b>
2891	buffy coat	M	36	1578	7624
2890	buffy coat	F	31	1316	13724
2810	SP	F	50	156	descartado
3060	SP	F	54	181	descartado
3308	buffy coat	F	37	1700	13000

<b>Líneas celulares</b>		
	<b>Rearreglo BCR-ABL</b>	<b>No. Células</b>
<b>K562</b>	p210 b3a2	5x10 <sup>5</sup>
<b>HL-60</b>	negativas	5x10 <sup>5</sup>

Las muestras fueron transportadas a Brasil en hielo seco, descongeladas en hielo, centrifugadas con vacío, resuspendidas en 7µL de agua y cuantificadas por QUBIT 2.0 previo a la construcción de las bibliotecas (Tabla 4.2).

Tabla 4.2. Resultado de la cuantificación de ARN mediante QUBIT en las muestras utilizadas para la construcción de las bibliotecas para secuenciación.

	<b>ARN (ng/µL)</b>
<b>Pool 1 HSC</b>	Fuera de rango (<20ng/µl)
<b>Pool 1 LSC</b>	
<b>Pool 2 HSC</b>	
<b>Pool 2 LSC</b>	
<b>Pool DS HSC</b>	
<b>Pool DS progenitores</b>	30
<b>K562</b>	500
<b>HL60</b>	360

El protocolo realizado se detalla en la sección II.6.3 de materiales y métodos. Algunas consideraciones especiales respecto al protocolo realizado se enumeran a continuación:

-El protocolo del kit *TruSeq Small RNA* de Illumina tiene un paso de purificación del producto de la biblioteca, en el cual se corren en paralelo, en un gel de poliacrilamida, los productos de reacción y marcadores de peso molecular que permiten la selección de la banda correspondiente a la biblioteca de microARNs (147pb) y piwiARNs (157pb) mediante la escisión del gel comprendido entre las bandas del marcador de peso molecular de 145 y 160pb (Fig. 4.8). En ninguna muestra se observó la banda esperada de 147pb. Sin embargo, dado que la baja masa del material de partida podría ocasionar que la sensibilidad de la visualización no sea suficiente, se recortó la sección del gel del tamaño hipotético esperado.

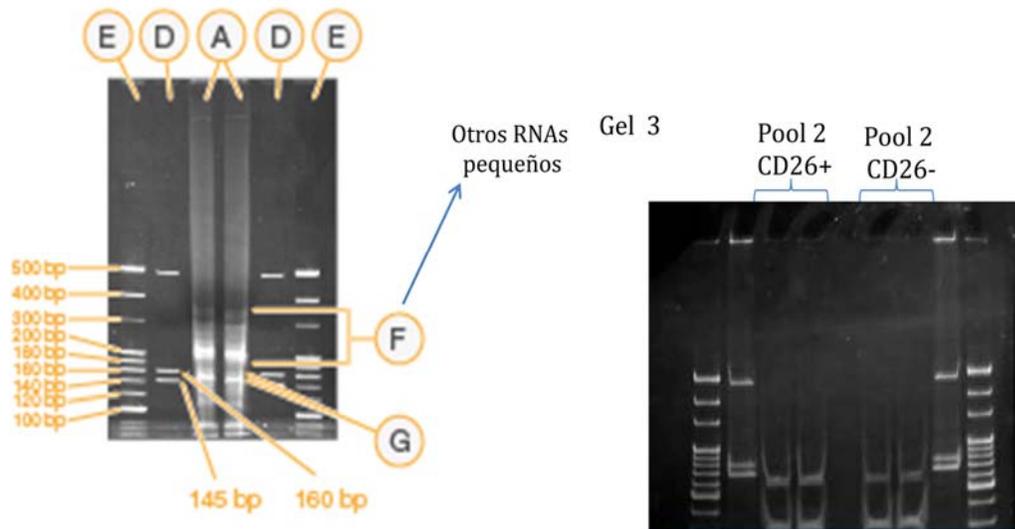


Fig. 4.8. Izquierda: marcadores de peso molecular y bandas esperadas según el fabricante. Los microARNs (G) están comprendidos entre las bandas de 145 y 160pb. Derecha: fotografía del gel obtenido al sembrar todo el volumen de las bibliotecas sintetizadas a partir de las muestras de pacientes correspondientes al pool 2. No se observa la banda del peso esperado. La banda inmediatamente inferior corresponde a dímeros de adaptadores.

#### IV. 4. Resultados de la secuenciación masiva del miRnoma de HSC y LSC.

##### IV. 4.1. Análisis de calidad de los datos

Se obtuvieron 200 millones de lecturas totales, con una densidad de clusters apropiada ( $600.000/\text{mm}^2$ ). Los datos generados fueron de buena calidad, con >90% de los datos con  $Q>30$  (Fig. 4.9).

El análisis de distribución de tamaños de las lecturas arrojó dos tipos de muestras: un grupo (1: pool 2 HSC, pool 2 LSC, HSC de DS, progenitores de DS) en el que se observó un pico alrededor de los 21-23 nucleótidos, correspondiente a los microARNs y lecturas de tamaños variables, que incluirían otros ARNs pequeños, y otro grupo (2: pool 1 HSC, pool 1 LSC, HL-60, K562) en el que no se observó dicho pico, por lo tanto corresponde a muestras en las que ocurrió un problema durante la construcción de la biblioteca (Fig. 4.9). La evaluación de la calidad de secuencias en estas muestras evidencia que la mayor

parte de las lecturas corresponden a secuencias de los *primers* utilizados durante la construcción de la biblioteca.

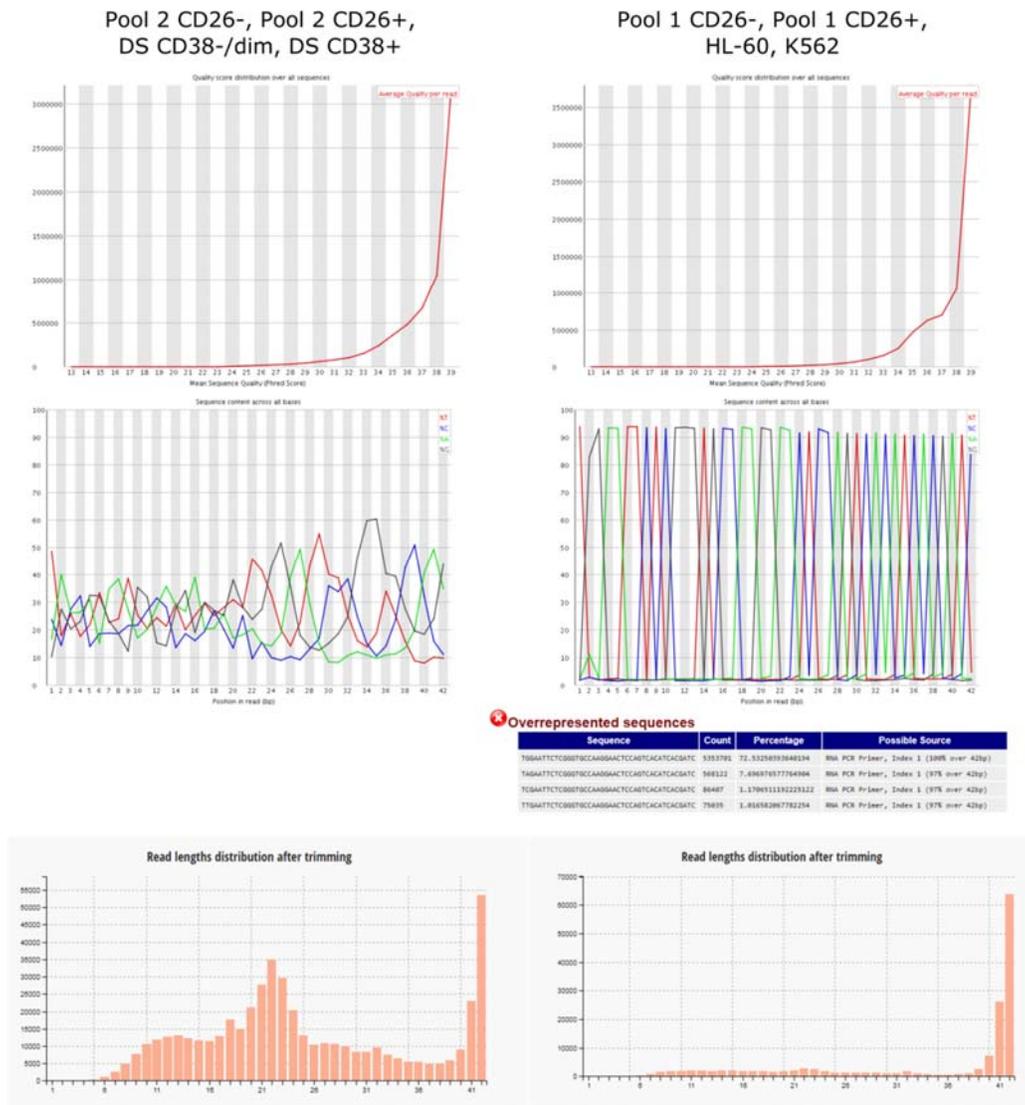


Fig. 4.9. Gráficas de calidad de secuencia representativas de los dos grupos observados. Grupo 1 (izquierda): se observa un pico de 21-23 nucleótidos correspondiente a los microARNs. Grupo 2 (derecha): no se observa el pico correspondiente a los microARNs y las gráficas de calidad evidencian que la mayor parte de las lecturas corresponden a secuencias de los *primers* utilizados durante la construcción de la biblioteca.

#### IV. 4.2. Análisis bioinformático para la identificación de microARNs

Se evaluaron distintas estrategias de análisis bioinformático de los datos. Los pasos principales llevados a cabo fueron:

- a) Remoción de lecturas de baja calidad y secuencias de adaptadores
- b) Identificación de microARNs conocidos

c) Análisis de expresión diferencial y predicción de potenciales ARNm blanco y vías asociadas

d) Descubrimiento de microARNs novel

### a) Remoción de lecturas de baja calidad y secuencias de adaptadores

A partir del análisis de calidad de las secuencias detallado en la sección anterior, sólo se prosiguió el análisis con las muestras de pacientes correspondientes al Pool 2 (HSC y LSC), y con las fracciones obtenidas a partir de DS. Se juntaron los datos provenientes de ambas calles (*lanes*) de la *flow cell* en un solo archivo. Se removieron las lecturas de baja calidad y las secuencias provenientes de adaptadores y *primers*.

### b) Identificación de microARNs conocidos

Se utilizó el programa Chimira<sup>69</sup>, el cual realiza un alineamiento entre las secuencias introducidas y las secuencias de precursores presentes en miRBase correspondientes a una determinada especie. Luego infiere a partir de qué brazo del precursor se originó la molécula de microARN madura (5' o 3').

El análisis global inicial reveló una gran profundidad de secuenciación, detectándose más de 1000 microARNs diferentes en todas las muestras (si se consideran aquellos con al menos 1 lectura) o más de 600 microARNs diferentes (si se consideran aquellos con al menos 10 lecturas) (Fig. 4.10). Este resultado fue llamativo dado el bajo input de ARN inicial (nuestra estimación es de <140 ng por fracción CD38; dado el límite de detección del kit utilizado para cuantificación por QUBIT y el volumen total de muestra), y concuerda con un reporte reciente de que la cantidad de ARN utilizada para la construcción de la biblioteca (rango: 50-1000 ng) no afecta al número de microARNs detectados<sup>102</sup>. Más llamativo aún es que nuestros resultados indican una mayor sensibilidad (alrededor del doble de la reportada), debido probablemente a la gran profundidad de secuenciación alcanzada (Anexo I).

Por otro lado, el análisis de frecuencias relativas mostró que los diez microARNs más abundantes constituyeron un 52-65% del total de microARNs detectados, dependiendo de la muestra (Fig. 4.10). Un estudio del miRNoma comparativo entre tejidos humanos reportó niveles similares de predominancia de algunos microARNs según el tejido<sup>102</sup>, y en particular en sangre entera, se reportó miR-486 como el más abundante (>90%). En nuestro caso, mientras que las muestras de pacientes secuenciadas correspondieron a SP y MO, las de DS fueron exclusivamente de SP (tabla 4.1). Sin embargo, miR-486 se encontró con una abundancia máxima de 6% (en progenitores de DS), lo cual sugiere que el patrón de microARNs observado es particular de células primitivas y progenitoras hematopoyéticas.

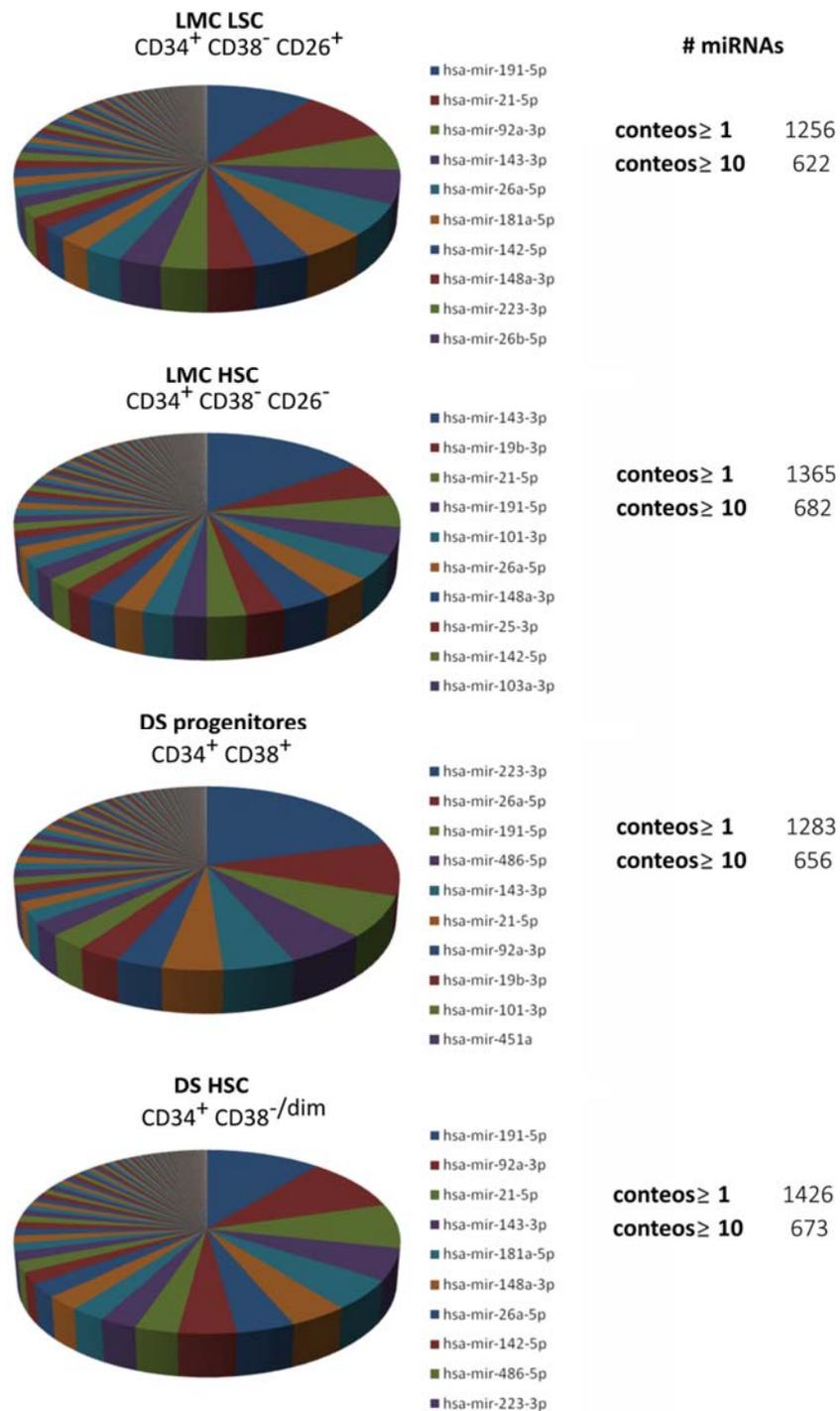


Fig. 4.10. Frecuencia de microARNs detectados en cada muestra.

### c) Análisis de expresión diferencial

El cálculo de expresión diferencial para los microARNs conocidos fue realizado con el programa GFOLD (*Generalized fold change for ranking differentially expressed genes from RNA-seq data*)<sup>70</sup>. Existen múltiples programas para realizar este tipo de análisis con datos

de secuenciación masiva. Sin embargo, dada la limitación de no tener réplicas biológicas, se buscó utilizar algún método que pueda tener dicha limitación en consideración. GFOLD está reportado como especialmente útil para el caso de no haber réplicas, ya que considera la distribución posterior del log *fold change*, calculando un cambio relativo en los niveles de expresión. Para ello tiene en cuenta los niveles absolutos detectados de cada microARN, ya que el *fold change* para genes con bajo número de lecturas es menos confiable que para genes de alta expresión.

Los datos correspondientes al análisis de expresión diferencial fueron filtrados de acuerdo a las siguientes preguntas:

- A) ¿Qué conjunto de microARNs caracteriza a las LSC?
- B) ¿Qué diferencias presentan las HSC de LMC vs. las HSC de dadores sanos?

A) ¿Qué conjunto de microARNs caracteriza a las LSC?

Se consideraron todos aquellos microARNs diferencialmente expresados (usando como línea de corte  $GFOLD \geq |1|$ ) en la comparación LSC vs HSC de pacientes con LMC, y se realizó la intersección con aquellos diferencialmente expresados en la comparación LSC vs. HSC de DS (Fig. 4.11). La intersección resultó en un listado inicial de 16 microARNs, siendo 8 subexpresados y 8 sobreexpresados en la fracción de LSC en comparación con las HSC de pacientes (Tabla 4.3).

A) ¿Qué conjunto de microARNs caracteriza a las LSC?

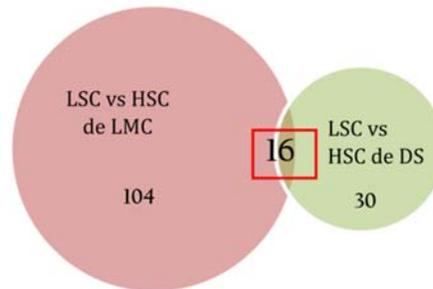


Fig. 4.11. Número de microARNs diferencialmente expresados ( $GFOLD \geq |1|$ ) en las comparaciones consideradas para la pregunta A). El conjunto seleccionado se muestra recuadrado.

Tabla 4.3. Valores de GFOLD y número de lecturas (relativizadas por DESeq) para los microARNs correspondientes al listado seleccionado como microARNs característicos de la fracción de LSC (pregunta A).

	GFOLD		Conteos relativizados (DESeq)			
	HSC LMC vs LSC	HSC DS vs LSC	LSC (LMC)	HSC (LMC)	HSC (DS)	Progenitores (DS)
hsa-mir-92b-3p	-3,6	1,6	2062	207	5951	505
hsa-mir-196a-5p	-3,6	-2,7	135	6	13	22
hsa-mir-126-5p	-2,0	1,1	18667	5343	44262	136969
hsa-mir-125a-5p	-1,9	-1,0	6225	2114	3045	1323
hsa-mir-2355-5p	-1,5	-1,0	928	385	444	602
hsa-mir-99b-5p	-1,2	-1,7	3907	2214	1207	4199
hsa-mir-411-5p	-1,2	3,6	36	9	749	132
hsa-mir-10a-5p	-1,1	1,1	28770	19727	68583	23278
hsa-mir-708-5p	1,6	2,5	4	46	42	4
hsa-mir-431-5p	1,6	1,9	8	112	89	27
hsa-mir-134-5p	1,9	2,5	1	64	62	87
hsa-mir-485-3p	2,2	2,7	1	53	46	41
hsa-mir-409-3p	2,6	3,8	46	595	902	663
hsa-mir-323b-3p	2,8	2,2	3	140	60	36
hsa-mir-432-5p	2,9	2,4	0	105	65	26
hsa-mir-382-5p	3,7	2,9	3	130	52	38

El GFOLD puede considerarse como un  $\text{Log}_2(\text{Fold change})$  robusto,  $\text{GFOLD} < 0 =$  sobreexpresados en LSC;  $\text{GFOLD} > 0 =$  subexpresados en LSC.

B) ¿Qué diferencias presentan las HSC de LMC vs. las HSC de DS?

Bajo el supuesto de que las HSC obtenidas a partir de muestras de pacientes con LMC **no** son equivalentes a las HSC de DS, propusimos la hipótesis de que las presentes en pacientes albergan alteraciones genéticas y/o epigenéticas relacionadas con el proceso leucemogénico, ya que se podrían ver afectadas por la coexistencia con LSC en el nicho de la MO.

B) ¿Qué diferencias presentan las HSC de LMC vs las HSC de dadores sanos?

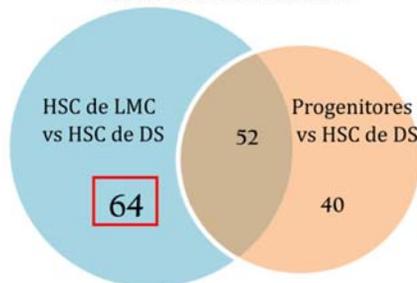


Fig. 4.12. Número de microARNs diferencialmente expresados ( $\text{GFOLD} \geq |1|$ ) en las comparaciones consideradas para la pregunta B). El conjunto seleccionado se muestra recuadrado.

Se compararon los microARNs diferencialmente expresados en esta fracción con respecto a los DS (usando como línea de corte  $\text{GFOLD} \geq |1|$ ). Dado que el número obtenido fue muy grande (116, Fig. 4.12), se excluyeron aquellos microARNs que resultaron también

diferencialmente expresados entre la fracción de progenitores y HSC de DS (52, Fig. 4.12). Ésto fue realizado dado que la población obtenida a partir de DS contenía, además de las CD38<sup>+</sup>, la población CD38<sup>dim</sup>, que incluye células más diferenciadas, y por lo tanto las diferencias observadas podrían deberse a la presencia de esta población adicional. Se encontraron 64 microARNs candidatos, de los cuales 17 (26%) resultaron sobreexpresados en HSC de LMC con respecto a las HSC de DS, y los 47 restantes (74%) se encontraron subexpresados. Dado que el listado de microARNs era muy extenso, se seleccionaron aquellos con  $GFOLD \geq |2|$  en la comparación HSC de DS vs. HSC de LMC (16 microARNs) (Tabla 4.4).

Tabla 4.4. Valores de GFOLD y número de lecturas (relativizadas por DESeq) para los microARNs correspondientes al listado seleccionado como microARNs diferencialmente expresados en HSC de LMC vs. HSC de DS (pregunta B).

	GFOLD	Conteos relativizados (DESeq)			
	HSC DS vs HSC LMC	LSC (LMC)	HSC (LMC)	HSC (DS)	Progenitores (DS)
hsa-mir-1304-3p	-3,6	563	1179	765	774
hsa-mir-629-5p	-2,3	207	2841	421	875
hsa-mir-629-3p	-2,3	234	1431	167	173
hsa-mir-362-5p	-2,0	984	7186	1204	1316
hsa-mir-20a-5p	2,0	9183	4054	16670	17420
hsa-mir-195-5p	2,1	75	16	104	60
hsa-mir-654-3p	2,4	23	67	516	373
hsa-mir-224-5p	2,5	20	2	36	52
hsa-mir-144-3p	2,6	1548	415	3244	3523
hsa-mir-181c-5p	2,7	19853	2315	12831	15289
hsa-let-7c-5p	2,7	616	236	896	740
hsa-let-7f-5p	2,9	159895	23477	139037	144680
hsa-mir-548o-3p	3,0	176	11	157	106
hsa-let-7a-5p	3,0	142802	17636	120313	111866
hsa-mir-182-5p	3,5	4267	4567	41803	33988
hsa-mir-183-5p	3,8	446	328	4196	3261

El GFOLD puede considerarse como un  $\text{Log}_2(\text{Fold change})$  robusto,  $GFOLD < 0$ : sobreexpresados en HSC de LMC;  $GFOLD > 0$ : subexpresados en HSC de LMC

#### d) Descubrimiento de microARNs *novels*

El análisis de microARNs *novels* se realizó mediante la herramienta miRDeep2. Esta herramienta utiliza un modelo probabilístico basado en el mecanismo de biogénesis de los microARNs, tal que evalúa la compatibilidad de la posición y frecuencia del ARN secuenciado con la estructura del precursor hipotético. Dado que en general existe una diferencia en la cantidad de lecturas provenientes de un microARN maduro (el que actuará como guía en el complejo RISC) vs. el ARN correspondiente al brazo opuesto del precursor, el algoritmo considera como potencial microARN a aquel que cumpla con una cierta diferencia en abundancia entre ambas moléculas (Fig. 4.13).

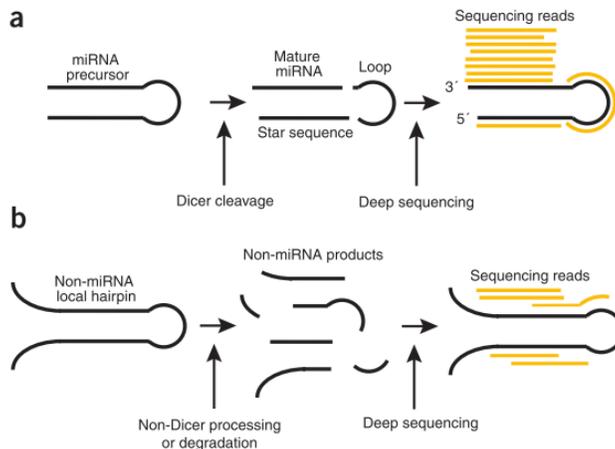


Fig. 4.13. Modelo probabilístico en el que se basa miRDeep2 para la predicción de microARNs novel. En el caso de un microARN de biogénesis canónica (a), una de las hebras del precursor será más estable que la otra y por lo tanto se reflejará en una mayor cantidad de lecturas en la secuenciación masiva. En el caso de un ARN pequeño, no microARN (b), por ej. resultante de la degradación de un ARN de mayor tamaño, a pesar de que pueda predecirse una estructura de hairpin de un precursor hipotético, no habría diferencia en la abundancia entre las hebras. Extraído de <sup>71</sup>.

Se seleccionaron aquellas predicciones con score  $\geq 1$  (ver Anexo I) y se realizó una intersección entre las listas de microARNs *novels* obtenidas por miRDeep2 para cada muestra. De éstas, 3 secuencias estaban presentes en todas las muestras:

Novel_miR_1	5' cagttgttcaaccagttac 3'
Novel_miR_2	5' atggtgtacatgatgtttgat 3'
Novel_miR_3	5' aaggtagatagaacaggtcttg 3'

Se evaluó el número de lecturas en los resultados de secuenciación, y sólo el novel\_miR\_3 era abundante en todas las muestras (ver Anexo I).

#### IV.5. Validación de los resultados de secuenciación masiva por RT-qPCR.

Las tecnologías de secuenciación masiva han permitido el desarrollo de nuevos métodos para el análisis global de la expresión génica. En comparación con la tecnología de *microarrays* de expresión, el análisis del transcriptoma por *RNA-seq* ha demostrado un mayor rango dinámico, un menor límite de detección, y permitido el descubrimiento de transcritos *novel* y variantes de *splicing*<sup>103</sup>. Respecto a la caracterización de la fracción de ARNs pequeños por secuenciación masiva, los primeros trabajos reportaron un aumento tanto en la variedad de secuencias detectadas<sup>104</sup>, como en la proporción de microARNs expresados diferencialmente con respecto a análisis previos<sup>105</sup>. Sin embargo, estas técnicas no son libres de limitaciones y sesgos propios. En particular, las fuentes de variabilidad que pueden potencialmente afectar los resultados de secuenciación masiva de ARNs pequeños son:

- la pérdida selectiva de microARNs particulares debido al método utilizado para extraer ARN y la cantidad de células de partida<sup>106</sup>;

- la amplificación selectiva de microARNs particulares según el protocolo utilizado durante la construcción de la biblioteca <sup>107-109</sup>;

Ésto implica que luego de un experimento de descubrimiento en una plataforma de caracterización masiva, como *microarrays* o secuenciación masiva, es mandatorio realizar una validación técnica de los resultados obtenidos. En nuestro caso, nos propusimos realizar no sólo una validación técnica utilizando la técnica de RT-qPCR, sino también biológica, evaluando nuevas muestras de pacientes y dadores sanos. Adicionalmente, la RT-qPCR requiere una menor cantidad de ARN inicial que la secuenciación masiva, y por lo tanto no fue necesario combinar muestras provenientes de distintos pacientes, lo que nos permitió evaluar la variabilidad intra-grupo.

Los microARNs evaluados fueron seleccionados a partir del análisis de expresión diferencial, como aquellos sobreexpresados en LSC de LMC (pregunta A). Se consideraron los 5 microARNs con mayor valor de GFOLD en la comparación LSC vs. HSC de pacientes: hsa-miR-196a-5p, hsa-miR-92b-3p, hsa-miR-126-5p, hsa-miR-125a-5p, y hsa-miR-2355-5p. En forma paralela, se sumaron al panel:

-hsa-miR-10a-5p y hsa-miR-132-3p, surgidos de una búsqueda bibliográfica sobre microARNs de potencial relevancia en el compartimiento *stem* hematopoyético<sup>110,111</sup>;

- hsa-let-7a-5p, surgido de una búsqueda bioinformática de microARNs localizados en las regiones translocadas durante la formación del cromosoma Ph;

-hsa-miR-182-5p, surgido de la comparación B, como uno de los microARNs más subexpresados en HSC de LMC vs. HSC de DS (Tabla 4.4);

-Se adicionó al panel el novel-3.

#### IV.5.1 Puesta a punto de la detección por RT-qPCR multiplex

Para estimar la masa de ARN de cada fracción, se realizó una curva de calibración con ARN de la línea celular KU812, midiendo el gen de referencia snARN U6. Se observó una amplificación lineal (log ARN vs. Ct) en el rango de 10pg a 10ng de ARN por reacción de retrotranscripción (RT) (Fig. 4.14, A). Todas las muestras entraron en la curva de masas (rango: 50pg-6,7ng, mediana=0,3ng), exceptuando una muestra (6pg); además se observó una correlación negativa entre el número de eventos sorteados y el Ct para snARN U6 (Fig. 4.14, B), sugiriendo que la estimación de masa a través de la cuantificación de snARN U6 fue apropiada.

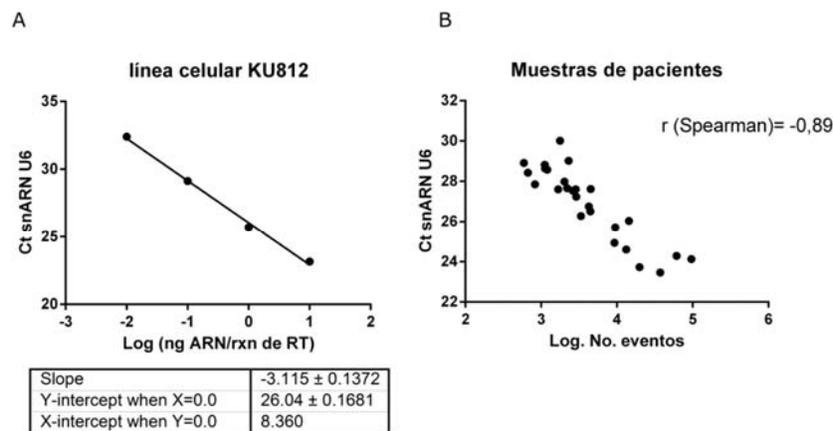


Fig. 4.14. RT-qPCR. A) Curva de calibración para snARN U6 utilizando distintas masas de ARN de la línea KU812 durante la reacción de RT. B) Correlación entre el valor de Ct para snARN U6, indicativo de la masa de ARN, y el número de eventos totales obtenidos por sorting a partir de cada fracción.

Se realizaron dos reacciones de RT en multiplex, las cuales incluyeron los siguientes microARNs:

<b>Multiplex 1 (M1)</b>	<b>Multiplex 2 (M2)</b>
hsa-miR-92b-3p	hsa-miR-132-3p
hsa-miR-125a-5p	hsa-let-7a-5p
hsa-miR-182-5p	hsa-miR-10a-5p
novel-3	hsa-miR-196a-5p
hsa-miR-2355-5p	
hsa-miR-126-5p	
<b>Control positivo: KU812</b>	<b>Control positivo: TF1a</b>

Las combinaciones de microARNs fueron seleccionadas tal que la eficiencia de amplificación en RT *multiplex* fuese similar a la obtenida en la condición de RT *singleplex*, y tal que sean detectables en el control positivo (línea celular). Para ello se realizaron curvas de masa para cada microARN (tabla 4.5). En el caso del hsa-miR-2355-5p, sólo se comparó el valor de Ct para el punto de mayor masa, ya que fue el único punto que amplificó en las líneas celulares evaluadas.

Tabla 4.5. Eficiencias de amplificación para cada microARN y snARN U6 en condiciones de RT *singleplex* o *multiplex* (M1 o M2 según corresponda). La fórmula utilizada para el cálculo de eficiencia se detalla en Materiales y métodos.

	<b>Eficiencia % (singleplex)</b>	<b>Eficiencia % (multiplex)</b>
<b>snARN U6</b>	Media=104%; Rango=89-120%	
<b>hsa-miR-92b-3p</b>	101	101
<b>hsa-miR-125a-5p</b>	94	97
<b>hsa-miR-182-5p</b>	103	102
<b>novel-3</b>	118	119
<b>Hsa-miR-2355-5p*</b>	Ct 36,25	Ct 37,16
<b>hsa-miR-126-5p</b>	107	103
<b>hsa-Let-7a-5p</b>	120	117
<b>hsa-miR-132-3p</b>	100	97
<b>hsa-miR-10a-5p</b>	108	106
<b>hsa-miR-196a-5p</b>	106	83

\*Comparación del Ct correspondiente al punto de mayor masa.

#### IV.5.2. Selección de pacientes

Para esta etapa del proyecto se procesaron 20 muestras de pacientes con LMC al diagnóstico. La fracción de células mononucleares fue analizada por citometría de flujo con el objetivo de seleccionar aquellas muestras con mayor proporción de células CD34<sup>+</sup>CD38<sup>-</sup> y por lo tanto mayor rendimiento (mediana[%CD38<sup>-</sup>]=1,5%, respecto a las CD34<sup>+</sup> totales), resultando en 10 muestras (mediana[%CD38<sup>-</sup>]=4,5%). De éstas, luego de realizar el *sorting*, se seleccionaron aquellas con mejor separación entre las poblaciones, quedando 6 para la etapa de validación por RT-qPCR.

Al igual que las muestras utilizadas para la secuenciación masiva, se observó que las HSC y LSC de pacientes con LMC difieren en su nivel de expresión de CD26, CD45, CD34, como así también en los parámetros asociados a tamaño celular (FSC) y complejidad interna (SSC) (Fig. 4.15). Se incluyó como tercera población a la fracción de células progenitoras

CD38+ de pacientes, a pesar de que la misma no había sido incluida en la etapa de secuenciación masiva. Además, se procesaron y sortearon 5 muestras de DS, de las cuales una fue excluida de los análisis posteriores por presentar mucha variabilidad en los resultados de RT-qPCR respecto a las demás. Las características de las muestras seleccionadas se detallan en la tabla 4.6.

*Tabla 4.6. Características de las muestras utilizadas para la etapa de validación. Se detallan el número de eventos obtenidos por sorting para cada fracción (para CD26- y CD26+ se indica el % respecto a la fracción CD34+CD38-), y el perfil de citometría.*

<b>LMC al Dx</b>						
<b>Código</b>	<b>Edad</b>	<b>Sexo</b>	<b>No. Ev. HSC (%)</b>	<b>No. Ev. LSC (%)</b>	<b>No. Ev. Progenitores</b>	<b>Perfil</b>
<b>N36 SP</b>	56	M	2857(51,3%)	2200(45%)	37272	1
<b>N47 SP</b>	58	M	1128(27,5%)	4504(62,5%)	96333	1
<b>N55 SP</b>	25	M	descartado	9276(89,1%)	19861	3
<b>N38 SP</b>	31	M	592(94,1%)	descartado	2300	2
<b>N46 SP</b>	38	F	descartado	3360(91,4%)	61428	3
<b>N56 SP</b>	55	M	4267(53,4%)	2910(36,6%)	291653	1

<b>DS</b>					
<b>Código</b>	<b>Edad</b>	<b>Sexo</b>	<b>No. Ev. HSC</b>	<b>No. Ev. Progenitores</b>	
<b>3984</b>	36	F	1117	1779	
<b>3308</b>	37	F	1682	2027	
<b>2771</b>	53	M	4460	9539	
<b>4532</b>	57	F	2314	13256	

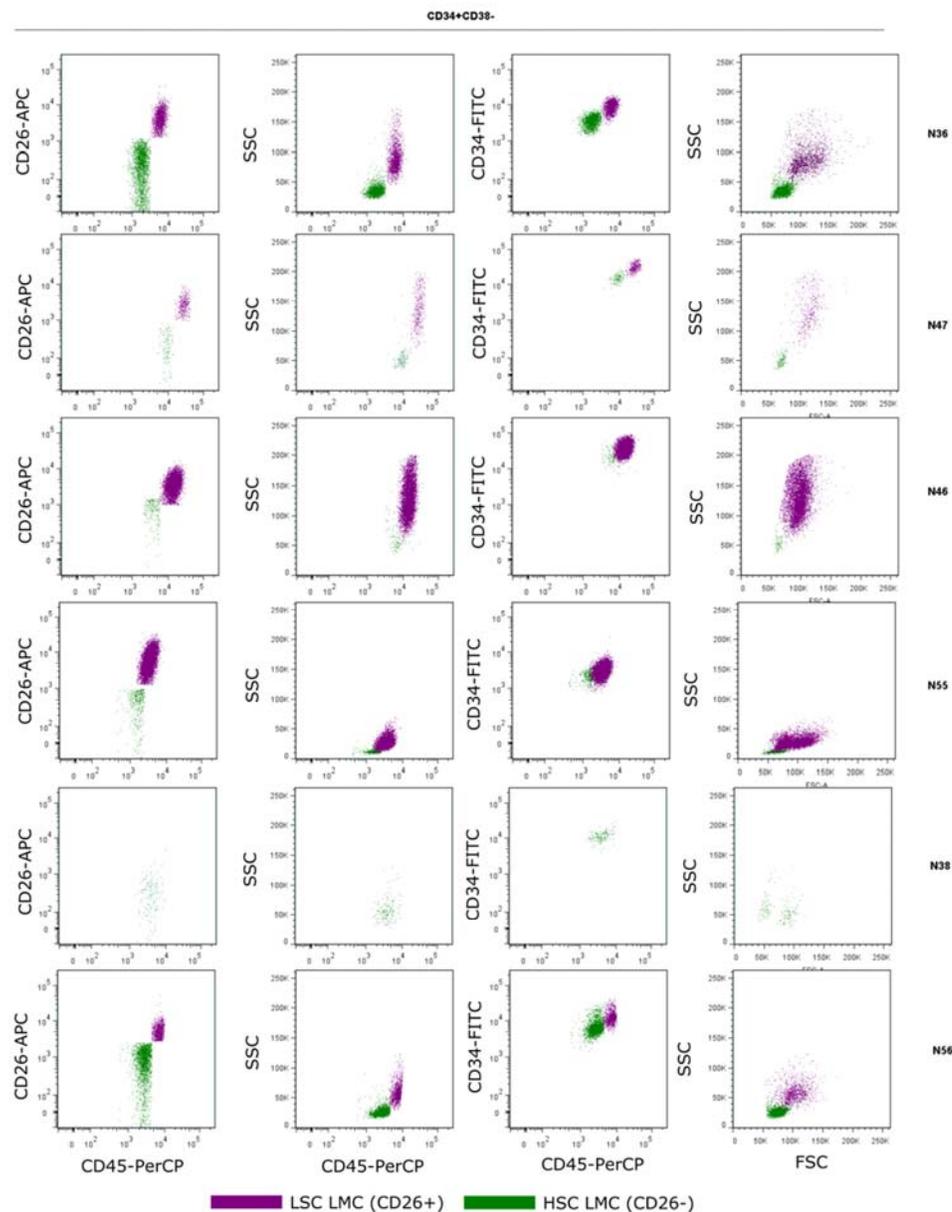


Fig. 4.15. Patrón de expresión de CD45, CD26, CD34 y FSC/SSC para las poblaciones de LSC y HSC de pacientes con LMC al diagnóstico utilizadas en la etapa de validación.

#### IV.5.3. Evaluación del perfil de citometría de flujo en pacientes con distinta respuesta al tratamiento con ITKs

Las diferencias observadas por citometría de flujo en la proporción de LSC y HSC en pacientes al diagnóstico nos llevó a preguntarnos si tendrían relación con la biología del tumor, el estado de avance de la enfermedad al momento del diagnóstico, y/o con la respuesta al tratamiento.

Se evaluó la frecuencia de cada perfil entre los pacientes, clasificados según:

- score Sokal (ver Introducción);
- RM al tratamiento con ITK 1<sup>ra</sup> línea a los 3 meses:  $\leq 10\%^{15}$  o  $>10\%^{15}$ ;

-RM al tratamiento con ITK 1<sup>ra</sup> línea a los 6 meses:  $<1\%^{IS}$  o  $\geq 1\%^{IS}$ .

La frecuencia de casos en cada grupo se muestra en la tabla 4.7. No se pudo realizar un análisis estadístico de tipo tablas de contingencia debido al bajo número de casos por categoría. Se observaron las siguientes tendencias (Fig. 4.16):

-una mayor proporción de casos con sokal alto en el perfil 3;

-ausencia de casos con  $RM > 10\%^{IS}$  a los 3 meses en el perfil 2;

-todos los casos con  $RM \geq 1\%^{IS}$  a los 6 meses presentaron un perfil 3.

*Tabla 4.7 Frecuencias de pacientes con perfil 1, 2 o 3 en la citometría de flujo al diagnóstico según score Sokal, RM a los 3 meses o 6 meses de tratamiento con ITK 1<sup>ra</sup> línea.*

Score Sokal	Perfil citometría de flujo			Total
	Perfil 1	Perfil 2	Perfil 3	
<b>alto</b>	1	1	3	5
<b>bajo</b>	4	2	1	7
<b>intermedio</b>	3	2	3	8
<b>Total</b>	8	5	7	20
Respuesta a ITK 1 <sup>a</sup> línea (3 meses)	Perfil 1	Perfil 2	Perfil 3	
<b>&gt;10%</b>	2	0	2	4
<b>≤10%</b>	6	5	5	16
<b>Total</b>	8	5	7	20
Respuesta a ITK 1 <sup>a</sup> línea (6 meses)	Perfil 1	Perfil 2	Perfil 3	
<b>≥1%</b>	0	0	3	3
<b>&lt;1%</b>	5	4	3	12
<b>Total</b>	5	4	6	15

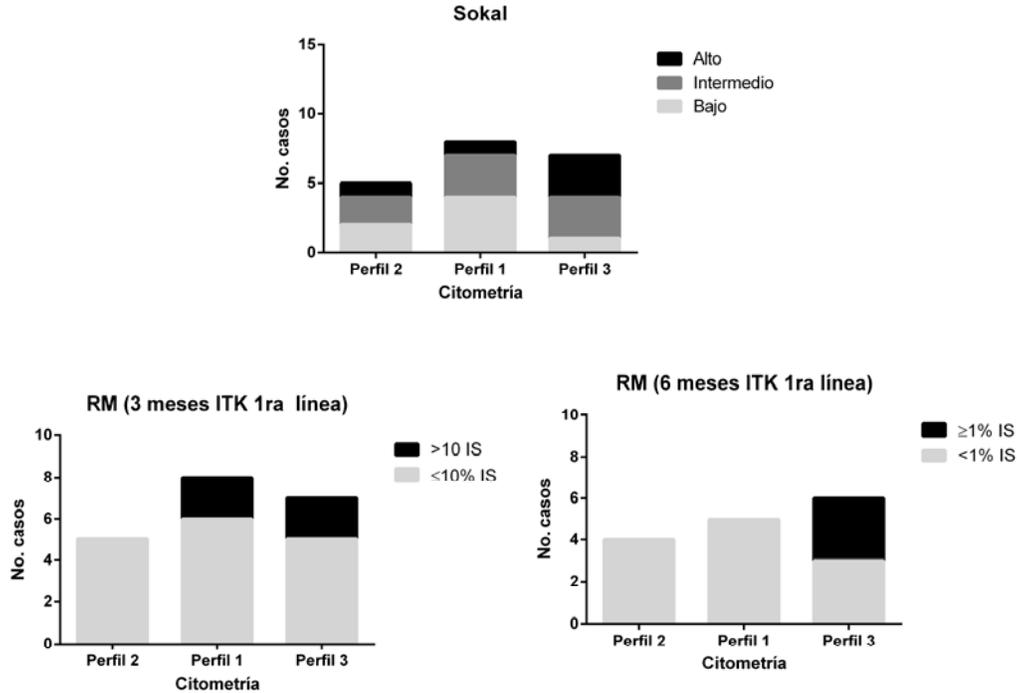


Fig. 4.16. Distribución de casos según perfil de citometría de flujo al diagnóstico, para los pacientes clasificados según score Sokal, y según su RM a los 3 y 6 meses de tratamiento con ITK en 1<sup>ra</sup> línea.

#### IV.5.4. Cuantificación de los niveles de microARNs por RT-qPCR.

El microARN hsa-miR-2355-5p no pudo ser detectado en la mayoría de las muestras, mientras que hsa-miR-182-5p sólo pudo ser detectado en las fracciones CD38<sup>+</sup> de LMC. Por lo tanto, ambos microARNs han sido excluidos de los análisis posteriores.

##### IV.5.4.1. Análisis de expresión diferencial

Los niveles de microARNs fueron comparados entre fracciones mediante el cálculo de la diferencia en el Ct entre el microARN correspondiente y snARN U6 (dCt) (Fig. 4.17). Se utilizó dCt como variable para el análisis estadístico (ver materiales y métodos). La prueba global fue estadísticamente significativa para hsa-miR-196a-5p, hsa-miR-10a-5p, hsa-miR-125a-5p, hsa-miR-92b-3p y hsa-miR-126-5p ( $p < 0,05$ , ANOVA); no se detectaron diferencias estadísticamente significativas en la prueba global para hsa-miR-132-3p, hsa-let-7a-5p y novel-3 ( $p > 0,05$ , ANOVA). El análisis detallado de las diferencias observadas se describirá en la siguiente sección.

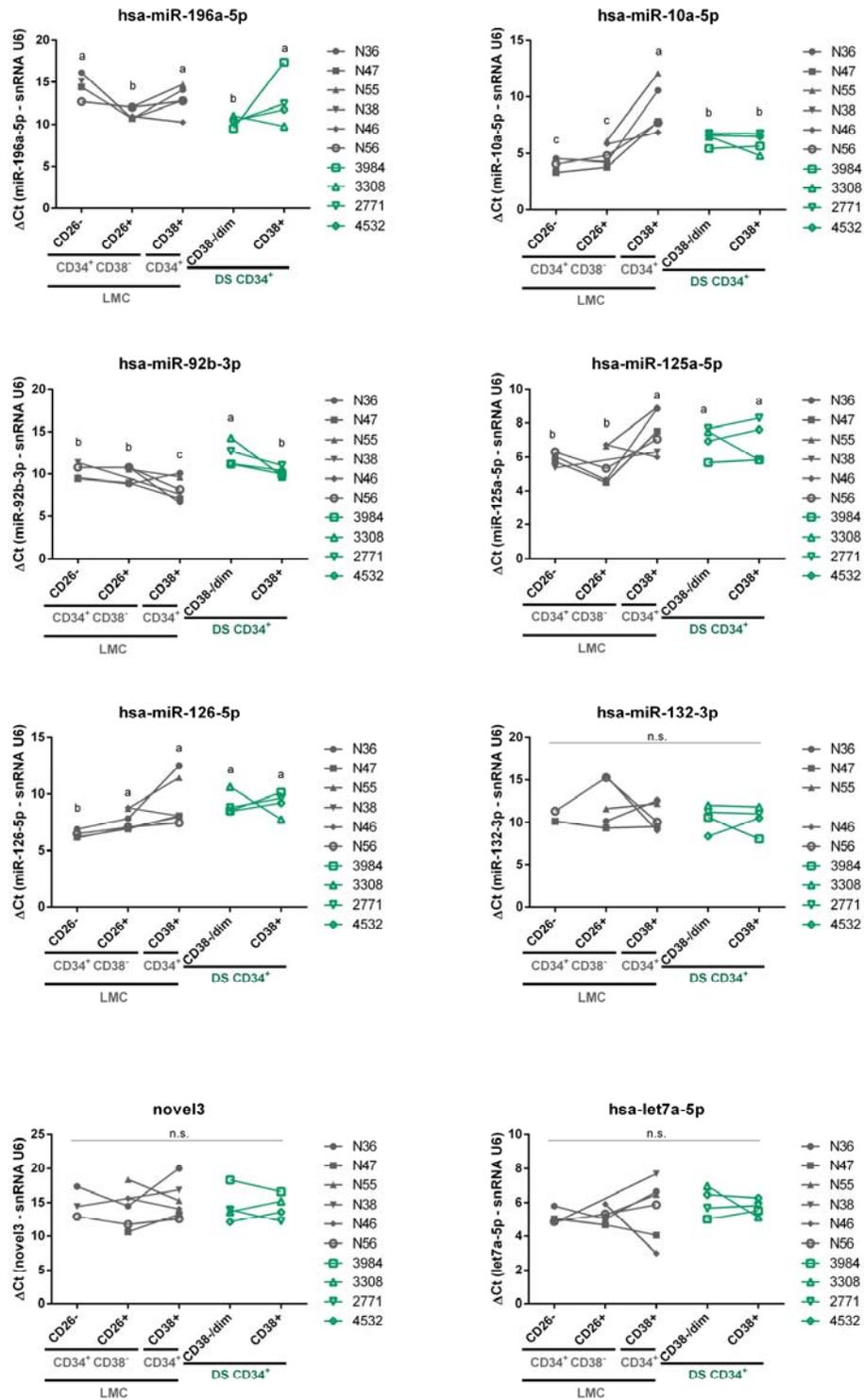


Fig. 4.17. Comparación del  $\Delta Ct$  entre los microARNs indicados y el gen de referencia snARN U6 para cada fracción. Los datos son el promedio de duplicados. Se indican los datos provenientes de los mismos pacientes o DS mediante líneas conectoras. Letras distintas indican diferencias estadísticamente significativas (ANOVA, comparación a posteriori, método DCG,  $\alpha$  global = 0,05)

A diferencia de lo ocurrido durante la secuenciación masiva, en este caso se pudo evaluar la variabilidad entre individuos dentro de cada grupo. Los pacientes o dadores mostraron, en su mayoría, tendencias similares al comparar entre poblaciones celulares, con excepciones para todos los microARNs evaluados. A su vez, la variabilidad intra-grupo fue mayor para las células progenitoras (CD38<sup>+</sup>) de pacientes, sugiriendo que se trata de una fracción más heterogénea (Fig. 4.18).

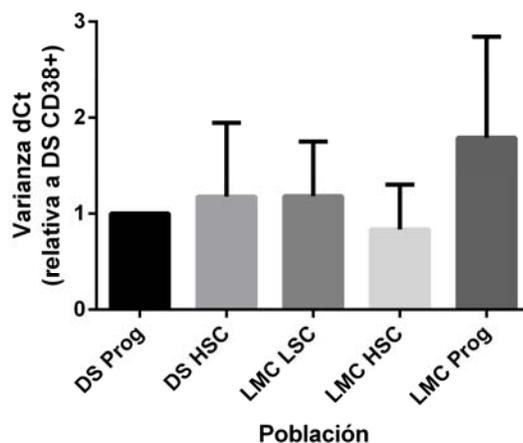


Fig. 4.18. Estimación de la varianza promedio por grupo para todos los microARNs. Se realizó una relativización a la varianza del dCt del grupo de progenitores de DS para cada microARN, y luego se promedió dicho valor para todos los microARNs evaluados.

#### IV.5.4.2. Correlación entre secuenciación masiva y RT-qPCR

Se comparó el  $\log_2$  de las veces de cambio (FC: *fold change*) para cada microARN entre ambas técnicas. Para los datos obtenidos por RT-qPCR, se utilizó el método  $2^{-\Delta\Delta Ct}$ ; en el caso de la secuenciación masiva, los datos utilizados fueron los valores de GFOLD (Fig. 4.19). El novel-3 fue excluido de este análisis por no contar con un valor de GFOLD.

Se observó una correlación baja en la comparación de fracciones provenientes de pacientes, y nula en la comparación de fracciones provenientes de DS (Fig 4.19, Tabla 4.8). Los únicos microARNs que mantuvieron la misma tendencia que la detectada en la secuenciación masiva fueron aquellos seleccionados en la pregunta "A" (microARNs característicos de la fracción de LSC), a excepción de miR-126-5p, que mostró una tendencia exactamente opuesta entre ambas técnicas. Ninguno de los microARNs agregados al listado mantuvo la tendencia observada por secuenciación masiva, a excepción de let-7a-5p en la comparación LSC vs. HSC (DS) (Tabla 4.8)

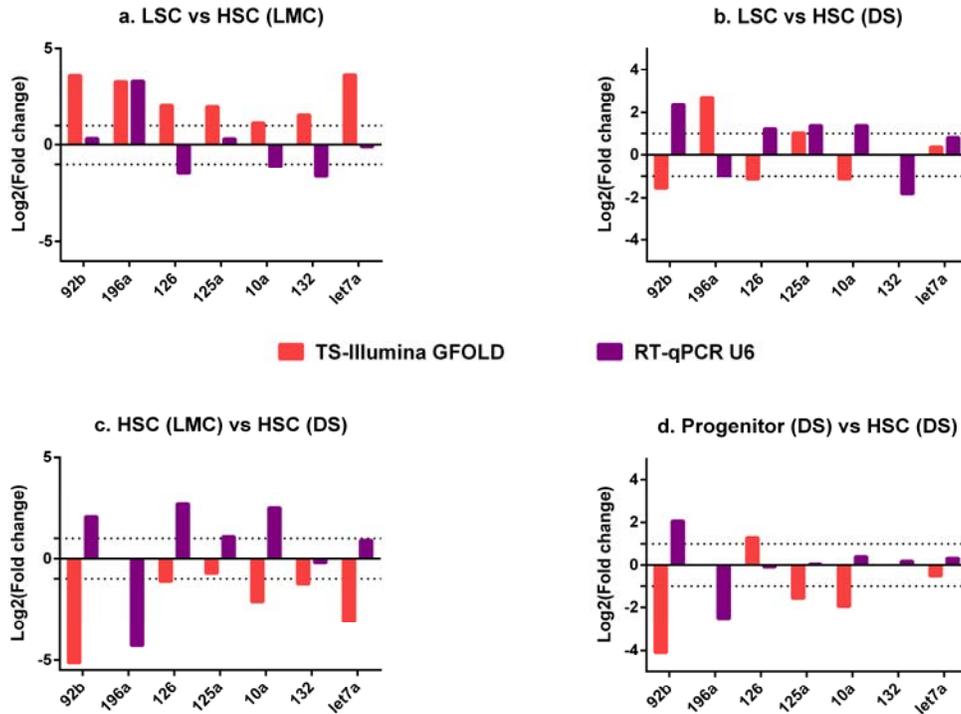


Fig. 4.19.  $\text{Log}_2(\text{FC})$  de los microARNs cuantificados por secuenciación masiva (TS-Illumina, método GFOLD) o por RT-qPCR (relativización por snARN U6), comparando las distintas fracciones entre sí. La línea punteada indica  $\text{Log}_2(\text{FC})=1$ .

Tabla 4.8. Resumen de la comparación de  $\text{Log}_2(\text{FC})$  entre la secuenciación masiva y la RT-qPCR. El p-valor hace referencia a la significancia estadística de las comparaciones a posteriori de los valores de dCt obtenidos por RT-qPCR.

	=tendencia p<0,05	=tendencia p>0,05	≠tendencia p<0,05	≠tendencia p>0,05
a. LSC vs HSC (LMC)	196a-5p	92b-3p 125a-5p	126-5p	10a-5p let-7a-5p 132-3p
b. LSC vs HSC (DS)	125a-5p	hsa-let-7a-5p	10a-5p 92b-3p	196a-5p 126-5p 132-3p
c. HSC (LMC) vs. HSC (DS)			125a-5p 10a-5p 92b-3p 196a-5p 126-5p	let-7a-5p 132-3p
d. Progenitores (DS) vs. HSC (DS)			196a-5p 92b-3p	125a-5p 10a-5p 126-5p let-7a-5p 132-3p

En conclusión, las diferencias en los niveles de microARNs que pudimos detectar con ambas metodologías y ambas cohortes de pacientes y dadores sanos fueron:

-en pacientes con LMC, aumento de los niveles de **hsa-miR-196a-5p** en la fracción de LSC (CD26+) en comparación con HSC (CD26-);

-aumento de los niveles de **hsa-miR-125a-5p** en la fracción de LMC LSC (CD26+) en comparación con la fracción de HSC (CD38-/dim) de dadores sanos.

Por otro lado, la evaluación de la fracción de células progenitoras CD38+ en pacientes con LMC, realizada sólo por RT-qPCR, en una cohorte de pacientes y dadores sanos (Fig. 4.17), resultó en la detección de diferencias estadísticamente significativas en los niveles de **hsa-miR-92b-3p** (aumentados en LMC CD38+) y **hsa-miR-10a-5p** (disminuidos en progenitores de LMC) con respecto a las restantes, tanto de pacientes como dadores, mientras que para **hsa-miR-125a-5p**, se detectaron diferencias significativas (disminución) respecto a las poblaciones LMC HSC y LMC LSC. Los valores de Log<sub>2</sub>FC para esta fracción, comparada con todas las restantes se muestra en la Fig. 4.20.

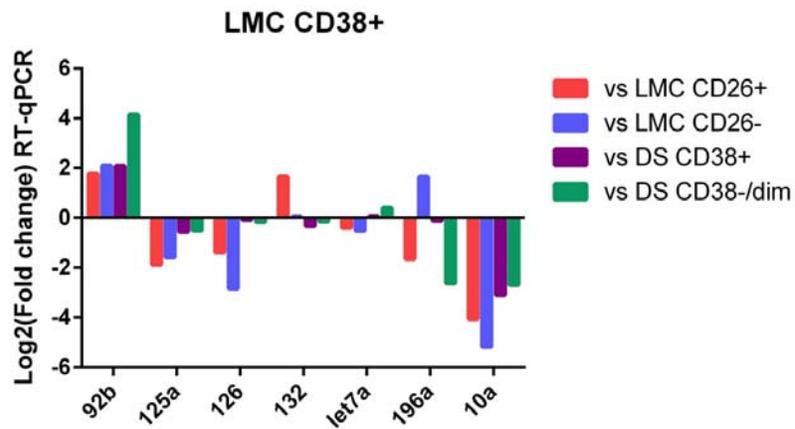


Fig. 4.20. Log<sub>2</sub>(FC) de los microARNs cuantificados por RT-qPCR (relativización por snARN U6) en la fracción de células progenitoras (CD38+) de LMC vs. las fracciones restantes.

Finalmente, con los datos obtenidos por RT-qPCR se construyó un mapa de color para indicar los niveles relativos de los microARNs cuya prueba global dio significativa, entre las fracciones correspondientes a pacientes (Fig. 4.21). Dicha representación resalta, en pacientes con LMC al diagnóstico, la presencia de hsa-miR-10a-5p y hsa-miR-125a, y la ausencia de hsa-miR-92b-3p en la fracción de células primitivas. No se tuvieron en cuenta los resultados obtenidos con muestras de DS dada la falta de reproducibilidad evidenciada al comparar los resultados con los de secuenciación masiva.

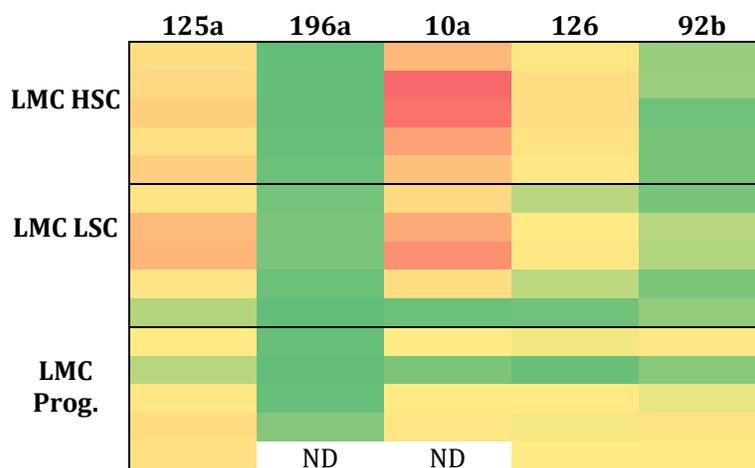


Fig. 4.21. Mapa de color de los niveles relativos de cada microARN a partir de los resultados de RT-qPCR ( $2^{-dCt}$ ), respecto a los niveles de todos los microARNs en todas las muestras. La escala va desde el rojo (mayores niveles relativos) a verde (menores niveles relativos). Amarillo indica valores intermedios.

### Novel-3

La cuantificación del novel-3 por RT-qPCR resultó en niveles muy bajos ( $Ct > 38$ ), y no se detectaron diferencias significativas entre las fracciones estudiadas (Fig. 4.18).

### IV.5.5. Análisis multivariado

Con el objetivo de detectar patrones globales de desregulación de los niveles de microARNs, se evaluó la correlación en la abundancia de los microARNs estudiados en la etapa de validación, utilizando como estimador de la abundancia la variable  $dCt = (Ct \text{ microARN} - Ct \text{ snARN U6})$  (Tabla 4.9). Se observó una correlación positiva ( $r > 0,5$ ) entre miR-125a-5p, miR-126-5p y miR-10a-5p.

Tabla 4.9. En color celeste se indican los coeficientes de correlación de Pearson para cada par de variables, y en verde las probabilidades asociadas a la prueba de hipótesis de correlación nula.

	126	92b	125a	novel3	132	let7a	10a	196a
126	1	0,25	5,10E-05	0,05	0,53	0,02	5,50E-06	0,73
92b	0,25	1	0,35	0,95	0,11	0,07	0,66	0,59
125a	0,74	0,2	1	0,47	0,33	0,02	5,70E-07	0,57
novel3	0,42	-0,02	0,16	1	0,58	0,29	0,26	0,2
132	0,15	0,37	0,23	0,14	1	0,19	0,57	0,67
let7a	0,49	0,39	0,47	2,40E-01	0,31	1	0,06	0,49
10a	0,81	-0,1	0,85	0,26	0,13	0,41	1	0,66
196a	0,08	-0,12	0,13	0,29	-0,1	0,16	0,1	1

Dado que los microARNs pueden formar parte de *clusters* genómicos, y que en este caso su expresión suele estar regulada de forma conjunta, se evaluó si los microARNs que mostraron correlación positiva pertenecían al mismo *cluster*; sin embargo, los tres microARNs están localizados en cromosomas diferentes.

Por otro lado, los miembros de una misma familia tienen secuencias conservadas y están relacionados evolutivamente, pudiendo estar relacionados funcionalmente, por ejemplo, regulando diferentes componentes de una misma vía de señalización (Tehler 2011). Encontramos que miR-125a y miR-10a pertenecen a la familia miR-10/100.

En el caso de microARNs codificados en el interior de otro gen (intragénicos), o muy cerca de otros genes, suele existir una relación funcional entre el microARN y el gen cercano. Por ejemplo, miR-10 y miR-196 han co-evolucionado con los *clusters* de genes Hox, esenciales durante el desarrollo como factores de transcripción maestros durante la formación del eje antero-posterior en animales bilaterales<sup>112</sup>.

Por último, la correlación en los niveles de microARNs podría deberse a una regulación conjunta, dada por los mismos factores (ej. río abajo del mismo factor de transcripción), ya sea directa o indirectamente. En este contexto, se encontró que miR-125a y miR-126 son mediadores de la respuesta a factores pro-angiogénicos, aunque las evidencias provienen de distintos modelos.

Estas características se resumen en la tabla 4.10.

Tabla 4.10. Características de los microARNs que mostraron correlación positiva entre las fracciones estudiadas durante la validación. CE: células endoteliales; LC: líneas celulares

	Hsa-miR-125a-5p	Hsa-miR-126-5p	Hsa-miR-10a-5p
<b>Secuencia microARN maduro</b>	ucccugagaccuuuaaccuguga	cauuuuuacuuuugg uacgcg	uaccuguagauccgaauuu gug
<b>Gen</b>	MIR125A	MIR126	MIR10A
<b>Loc. genómica</b>	19q13.41[+]	9q32.3[+]	17q21.32[-]
<b>Genes cercanos</b>	SPACAC (intragénico, exónico)	EGFL7 (intragénico, intrónico)	HOXB3 (intragénico, intrónico)
<b>Cluster</b>	miR-99b/let-7e	-	-
<b>Familia</b>	miR-125a, miR-10a, miR-125b-1, miR-125b-2, miR-10b, miR-99a, miR-99b, miR-100	-	miR-125a, miR-10a, miR-125b-1, miR-125b-2, miR-10b, miR-99a, miR-99b, miR-100
<b>Regulación transcripcional</b>	p53 (LC de cáncer de pulmón); NFkB (LC de cáncer de pulmón y hepatocarcinoma)	Coexpresión con su hospedador EGFL7 (expresión en CE)	TWIST1
<b>Mediadores positivos o negativos de la respuesta</b>	PDGFR	FGF y VEGF	?
<b>Referencias</b>	113,114	115	116

Finalmente, se analizaron las características de los microARNs restantes, en los que no se detectó una correlación positiva o negativa por RT-qPCR (Tabla 4.11). Los mismos se localizan en regiones genómicas distintas entre sí. Todos los microARNs pertenecen familias distintas. Es particular el caso de miR-196a, el cual está localizado en el *cluster* de genes HoxB, al igual que miR-10a, pero se encuentran unas 40kB de distancia.

Tabla 4.11. Características de los microARNs que no mostraron correlación entre las fracciones estudiadas durante la validación.

	<b>Hsa-miR-92b-3p</b>	<b>Hsa-miR-196a-5p</b>
<b>Secuencia microARN maduro</b>	uauugcacucgucggccucc	uagguaguuucauguuguugg
<b>Gen</b>	MIR92B	MIR196A1
<b>Loc. genómica</b>	1q22 [+]	17q21.32 [-]
<b>Genes cercanos</b>	No (Intergénico)	Hox B9
<b>Cluster</b>	-	-
<b>Familia</b>	miR-25, miR-92a-1, miR-92a-2, miR-92b, miR-363	miR-196a-1, miR-196a-2, miR-196b
	<b>Hsa-let-7a-5p</b>	<b>Hsa-miR-132-3p</b>
<b>Secuencia microARN maduro</b>	ugagguaguagguuguauaguu	uaacagucuacagccauggucg
<b>Gen</b>	MIRLET7A3	MIR132
<b>Loc. genómica</b>	22q.13.31	17p13.3[-]
<b>Genes cercanos</b>	MIRLET7BHG (lncRNA) (intragénico, exónico)	No (Intergénico)
<b>Cluster</b>	miR-4763, let-7b	miR-212
<b>Familia</b>	Let-7d, Let-7i, Let-7e, Let-7b, Let-7g, Let-7f-1, Let-7f-2, Let-7c, Let-7a-2, Let-7a-3, Let-7a-1, miR-98	miR-212, miR-132

#### IV.6. Conclusiones parciales y comentarios

A continuación, se discuten las principales conclusiones extraídas a partir de los resultados de esta segunda parte:

1) Se optimizó un protocolo para el aislamiento de fracciones de HSC y LSC provenientes de pacientes con LMC al diagnóstico de alta pureza.

Evaluamos múltiples condiciones para la obtención de perfiles de citometría que nos permitan la distinción clara entre las poblaciones, que sean de alta pureza (estimada por la proporción de ARNm de *BCR-ABL1*) y de las cuales se pueda obtener ARN conteniendo a la fracción de ARNs pequeños, de alta calidad. Las condiciones que mejor funcionaron en nuestro sistema fueron:

-La selección previa de pacientes con mayor proporción de células CD38<sup>+</sup> en la fracción de CD34<sup>+</sup>;

-La combinación de los marcadores de superficie CD45, CD34, CD38 y CD26, junto a los parámetros FSC y SSC para la separación de las poblaciones de HSC vs. LSC.

-La recolección de las células directamente en buffer de lisis para la extracción de ARN conteniendo a la fracción de ARNs pequeños, mediante un micrométodo que permite la elución en un volumen pequeño.

*2) Se logró realizar la secuenciación masiva a partir de <140ng de ARN inicial; sin embargo, el protocolo de construcción de la biblioteca tiene varias limitaciones.*

Obtuvimos una gran profundidad de secuenciación, aún habiendo partido de una cantidad de ARN menor a la reportada, lo cual nos permitió detectar microARNs de baja expresión (Fig. 4.10). Sin embargo, la principal limitación que encontramos durante la construcción de la biblioteca fue la selección de la banda de tamaño apropiado en un gel de poliacrilamida. Esta banda no se veía, mientras que una banda correspondiente al tamaño esperado por los dímeros de adaptadores sin inserto era evidente en todas las muestras. Según lo que pudimos consultar a los fabricantes del kit, la presencia de dichos dímeros de adaptadores en una alta proporción es una característica del método, que impacta negativamente ya que, de no ser eliminados, la mayor parte de las lecturas corresponden a la secuenciación de dichos adaptadores. Por otro lado, el hecho de no haber visto una banda definida correspondiente al producto de la biblioteca implicó la pérdida de la mitad de las muestras. Existen protocolos alternativos que buscan superar este problema mediante adaptadores modificados<sup>117</sup>.

*3) Se observó una tendencia a mayor frecuencia de casos con índice Sokal alto y  $RM \geq 1\%$  a los 6 meses en pacientes con perfil 3 por citometría de flujo al diagnóstico; también se observó una tendencia a menor frecuencia de casos con  $RM > 10\%$  a los 3 meses en pacientes con perfil 2.*

Dos trabajos reportan diferencias en la respuesta al tratamiento y parámetros al diagnóstico según el perfil de citometría al diagnóstico, aunque la distinción de las poblaciones no es exactamente la misma. En el trabajo de Janssen *et al* se reportó que aquellos pacientes que presentaban HSC residuales por citometría de flujo estaban asociados a menores índices Sokal y Euro, menor toxicidad por IM y mejor RM a IM que aquellos que no tenían HSC detectables al diagnóstico<sup>98</sup>. En el trabajo de Culen *et al* se reportó una asociación significativa entre el perfil de citometría y el número de glóbulos blancos al diagnóstico, y con la posibilidad de alcanzar una RM mayor a los 12 o 18 meses de tratamiento con Imatinib<sup>99</sup>.

*4) La correlación entre los resultados de la secuenciación masiva y la validación fue baja.*

El análisis que realizamos cuenta con varias limitaciones y fuentes de variabilidad que podrían explicar las diferencias observadas:

-La necesidad de combinar muestras provenientes de distintos pacientes para realizar la secuenciación masiva no nos permitió evaluar la variabilidad en el perfil de expresión de microARNs entre pacientes. Asimismo, la pérdida de uno de los pools

provenientes de pacientes implicó que los resultados obtenidos podrían ser particulares de las muestras evaluadas, y no ser representativos de todos los pacientes con LMC;

-El número de pacientes y DS fue bajo en ambos casos, debido a una limitación en el rendimiento de células primitivas de alta pureza y por la heterogeneidad presentada por los pacientes en los perfiles de citometría de flujo, siendo que en la mayoría de ellos no se distinguían claramente ambas poblaciones;

-En la secuenciación masiva se incluyeron muestras de MO y SP de pacientes, mientras que en la validación sólo se evaluaron muestras de SP.

A su vez, existen fuentes de variabilidad técnicas, intrínsecas a los métodos utilizados:

a) En particular para el kit Truseq de Illumina se ha reportado un sesgo significativo introducido por la secuencia de adaptadores que se utilizan durante la construcción de la biblioteca<sup>109</sup>, lo que ocasiona que algunos microARNs sean más eficientemente ligados que otros, y por lo tanto ese sesgo se mantiene en su representación en la biblioteca. Una de las posibles explicaciones es que la secuencia de los adaptadores influya en la eficiencia de ligación microARN-adaptador, debido al plegamiento diferencial exhibido por diferentes combinaciones de microARN-adaptador; para remediar este sesgo se han desarrollado métodos que utilizan adaptadores con bases al azar en los extremos<sup>118</sup>. Con el objetivo de evaluar la correlación en la abundancia de los microARNs entre ambas técnicas, se compararon los niveles relativizados por DESeq para los datos de secuenciación masiva, y diferencias en el Ct respecto a los niveles de snARN U6 para los datos de RT-qPCR. Se observó una correlación negativa, lo cual es esperable dado que el dCt es inversamente proporcional a la abundancia; a su vez, la correlación fue mayor para el subconjunto de muestras correspondientes a pacientes que para DS (Fig. 4.23).

b) Existen múltiples estudios que muestran el efecto de la normalización en la cuantificación por RT-qPCR. Mientras que algunos reportes han demostrado que los niveles de snARN U6 son menos estables respecto a otros microARNs propuestos como "normalizadores universales"<sup>119</sup>, otros han propuesto evaluar en cada sistema los microARNs más estables, y realizar una media geométrica de sus niveles como parámetro normalizador<sup>120</sup>. La molécula utilizada como normalizadora en nuestros experimentos podría estar afectando los resultados obtenidos, ya que la limitación de no poder cuantificar directamente la masa presente en las muestras impidió verificar la estabilidad del snARN U6 en las distintas condiciones evaluadas. Sin embargo, su correlación con el número de eventos sorteados sugiere que su utilización fue apropiada (Fig. 4.15, B). Luego de este análisis surgió el interés de buscar posibles microARNs normalizadores en nuestro sistema, a través de los resultados de la secuenciación masiva.

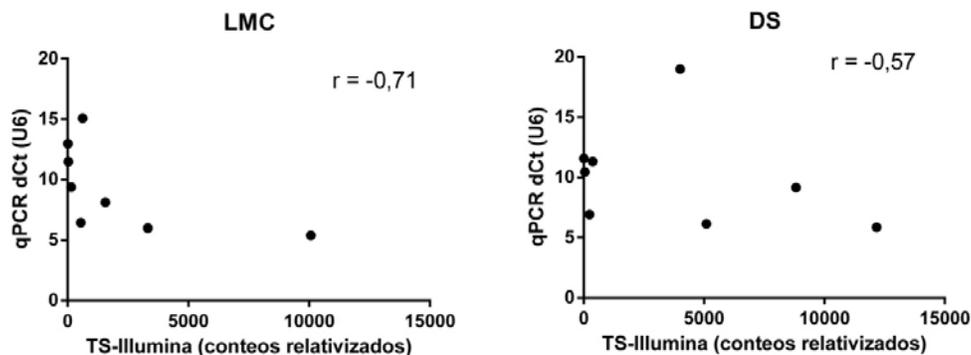


Fig. 4.23. Análisis de correlación (Spearman) para los datos de abundancia de los microARNs incluidos en la etapa de validación, entre la secuenciación masiva (kit TruSeq-Illumina) y RT-qPCR, tanto de pacientes (LMC) como donantes sanos (DS).

Por último, la comparación de los valores de  $\text{Log}_2\text{FC}$  evidenció una mejor correlación para aquellos microARNs desregulados entre las LSC y HSC de pacientes, mientras que los resultados obtenidos con muestras de DS resultaron opuestos. Desconocemos la causa de esta diferencia. Por un lado, es esperable que la variabilidad sea mayor entre DS que entre pacientes, ya que los primeros se definen por un criterio de exclusión ("sano": en ese momento no presenta una enfermedad evidente, pero desconocemos otras características relacionadas con su estado de salud), mientras que los segundos tienen un patrón común de desregulación del sistema hematopoyético relacionado al proceso leucemogénico. En este sentido, hubiese sido más apropiado evaluar más muestras de DS que de pacientes. Por otro lado, la selección de la fracción de HSC DS por sorting fue más difícil de estandarizar que la de HSC en pacientes, debido a la dificultad de fijar el umbral para definir a la población "dim" durante el proceso de sorting, lo cual pudo haber introducido variabilidad en los resultados.

5) La evaluación de los microARNs presentes en la fracción de progenitores de LMC (CD38+) por RT-qPCR evidenció que algunos de los microARNs seleccionados durante la secuenciación masiva también están desregulados en dicha población.

Éste fue el caso para miR-92b-3p, miR-10a-5p, y miR-125a-5p. Este patrón sugiere que dichos microARNs podrían estar asociados al proceso de diferenciación hematopoyética. En este contexto, miR-125a ha sido reportado como regulador de la capacidad de autorrenovación de las HSC, siendo que un aumento en los niveles del mismo causó un aumento en el pool de HSC-progenitores tempranos, a través de la regulación negativa de la apoptosis<sup>65</sup>. En el caso de miR-10a-5p, está reportada una disminución en sus niveles en la fracción de progenitores de LMC respecto a dadores sanos, y se demostró que una inhibición del mismo en líneas celulares conlleva a una mayor proliferación<sup>110</sup>. Por el contrario, en muestras de pacientes con Leucemia mieloide aguda portadores de la mutación en NPM1, se detectó un aumento en los niveles de miR-10a<sup>121</sup>. Asimismo, en otros tumores, por ejemplo, carcinoma hepático, cáncer de páncreas, cáncer de mama metastásico y leucemia linfática crónica de células B se ha detectado un aumento en los niveles de los miembros de la familia miR-10. En particular, para este microARN se ha descrito un mecanismo de acción no canónico como regulador positivo de la traducción de proteínas ribosomales, a través de la unión al 5'UTR de los ARNm de dichas proteínas<sup>63</sup>.

Por lo tanto, es posible que las consecuencias funcionales de los niveles alterados de miR-10a varíen según el sistema. Para miR-92b-3p, se reportó su participación en el control de la transición G1/S del ciclo celular en células humanas embrionarias<sup>122</sup>. No hay reportes de la asociación de miR-125a, miR-10a ni miR-92b a la fracción de células más primitivas en pacientes con LMC.

6) *El miR-126-5p mostró un patrón opuesto entre las fracciones LSC vs. HSC de pacientes.*

Este microARN ha sido extensamente estudiado por ser de relevancia en el sistema vascular. Se encuentra altamente conservado, y su localización genómica es el intrón 7 del gen EGFL7, el cual codifica para un péptido de expresión en células endoteliales. Ratones *knock-out* para miR-126 presentan deficiencias en la formación de vasos sanguíneos, hemorragias, y una fracción de los ratones sufre letalidad embrionaria<sup>115</sup>. En el sistema hematopoyético, miR-126 ha sido descrito como regulador de la capacidad *stem*, siendo que es de alta expresión en la fracción más primitiva del linaje hematopoyético humano (CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup>CD49f<sup>+</sup>); experimentos de *knock-down* en la fracción de células primitivas de cordón umbilical resultaron en un aumento de la proliferación *in vitro* e *in vivo*<sup>123</sup>. En LMC, se ha descrito una disminución en los niveles de miR-126 en la fracción de células primitivas de pacientes en comparación con DS, siendo sus niveles regulados por la actividad quinasa de BCR-ABL1. Dichos experimentos fueron realizados midiendo la forma madura correspondiente al brazo 3p (miR-126-3p), y no se informan los niveles de miR-126-5p<sup>124</sup>.

Sin embargo, dado que en este trabajo postulan que la regulación por parte de BCR-ABL1 implica un mecanismo dependiente del complejo que exporta microARNs del núcleo al citoplasma RAN-EXP5-RCC1, ocasionando una acumulación de los precursores en el núcleo y causando una disminución en los niveles del microARN maduro en el citoplasma, dicho mecanismo debería afectar también, en teoría al miR-126-5p. En nuestro caso, observamos una disminución en la fracción de LSC (BCR-ABL1<sup>+</sup>) vs. HSC (BCR-ABL<sup>-</sup>) de pacientes por RT-qPCR, sugiriendo que el patrón detectado por esta técnica, y no por secuenciación masiva, es el que refleja la regulación por parte de la actividad quinasa de BCR-ABL1. En este contexto, cabe preguntarse por qué se observó el patrón opuesto en la secuenciación masiva. Una posible explicación que planteamos (además de las fuentes de variabilidad generales ya descritas) es que, si las fracciones de pacientes utilizadas para la secuenciación masiva correspondían a muestras de MO y SP, mientras que aquellas utilizadas para la validación correspondían todas a SP, el patrón opuesto observado sugiere que los niveles de miR-126-5p podrían variar entre ambos sitios. En este sentido, el trabajo de Zhang *et al* demostró que las células endoteliales de la MO pueden transferir miR-126 a través de vesículas extracelulares a la fracción de células primitivas<sup>124</sup>ang. Para poner a punto esta hipótesis, sería de interés comparar los niveles de miR-126-3p y miR-126-5p en muestras de MO y SP de los mismos pacientes, en sus correspondientes fracciones LSC y HSC.

7) *Se detectó una correlación positiva entre los niveles de miR-125a-5p, miR-126-5p y miR-10a-5p por RT-qPCR*

Este resultado sugiere una posible regulación conjunta de miR-125a-5p, miR-126-5p y miR-10a-5p, ya que dicha correlación no pudo ser explicada por su localización genómica. miR-10a y miR-125a pertenecen a la misma familia, con lo cual es posible que formen parte de redes regulatorias relacionadas entre sí. En este sentido, en un modelo de cáncer de pulmón, miR-125a resultó regulado a nivel transcripcional por p53 y NFkB<sup>114</sup>, mientras que en un modelo de mielodisplasia, se observó que miR-10a puede mediar la sensibilidad a apoptosis inducida por TNF-a a través de la regulación de los niveles de p53 y de la actividad de la vía NFkB<sup>116</sup>. A su vez, está descrito que la fracción de LSC es especialmente dependiente de la actividad de p53<sup>95</sup>. Estos resultados sugieren que podría existir algún mecanismo común, en el que interactúen miR-125a, miR-10a y p53 en la fracción de LSC.

Por otro lado, la observación de que miR-126 y miR-125a son mediadores de la respuesta a factores pro-angiogénicos<sup>113,115</sup>, y que su expresión correlaciona en las muestras evaluadas, sugiere que ambos podrían participar de una misma vía de señalización en nuestro sistema.

## Capítulo V. Análisis *in silico* de microARNs

### V.1. Introducción

La identificación de microARNs diferencialmente expresados en las poblaciones celulares evaluadas permite al menos dos tipos de análisis:

- la evaluación de los perfiles de expresión en cada población;
- el estudio particular de cada microARN mediante ensayos funcionales.

Para ambos enfoques, utilizamos como punto de partida un análisis *in silico*, dada la gran cantidad de herramientas bioinformáticas y bases de datos disponibles para el estudio de microARNs. Las herramientas utilizadas consistieron en una combinación de los siguientes métodos:

- 1) Predicción de posibles blancos (ARNm)
- 2) Búsqueda de interacciones microARN-ARNm experimentalmente validadas
- 3) Cruce con listados de ARNm y proteínas desreguladas en células primitivas o progenitoras de LMC.

Los algoritmos bioinformáticos de predicción de posibles blancos (*targets*) de los microARNs se basan en el modelo canónico de acción de los microARNs, mediante la búsqueda de complementariedad entre la región semilla del microARN (nucleótidos 2 a 8) y la región 3'UTR de los ARNm, la predicción de estabilidad del complejo ARNm-microARN, y el grado de conservación del sitio blanco presente en el ARNm. Cada algoritmo otorga distinto peso a cada factor, resultando en *rankings* ligeramente diferentes. Sin embargo, estos algoritmos no tienen en cuenta aquellos mecanismos de acción de microARNs descriptos como no canónicos, como unión al 5'UTR<sup>63</sup>, o baja complementariedad entre la región semilla y el blanco, que es compensada por regiones complementarias hacia el extremo 3' del microARN<sup>58</sup>. Por otro lado, tienen una alta tasa de falsos positivos. Al agregar a las predicciones, los registros de interacciones experimentalmente validadas, estos rankings se amplían, ya que no todas las interacciones descriptas son canónicas.

Los experimentos de caracterización masiva (por ejemplo, secuenciación, *microarrays*, espectrometría de masa) requieren para su publicación que los datos sean almacenados en repositorios de acceso público o incluidos como anexos en las publicaciones mismas. La disponibilidad de dichos datos es de gran utilidad, ya que permite comparar los resultados propios con datos experimentales que tal vez no tengamos la posibilidad de generar. Aprovechar este tipo de datos permite, además, ahorrar tiempo y recursos. En nuestro caso, utilizamos dos listados de interés:

- un listado de ARNm proveniente de una caracterización del transcriptoma a nivel de células únicas, comparando la fracción de células primitivas Lin-CD34<sup>+</sup>CD38<sup>-</sup> (que expresaban el ARNm de *BCRABL*) vs. Lin-CD34<sup>+</sup>CD38<sup>-</sup> (que no expresaban el ARNm de *BCRABL*) de pacientes con LMC al diagnóstico<sup>96</sup>;
- un listado de proteínas proveniente de la caracterización del proteoma por espectrometría de masa, comparando la fracción CD34<sup>+</sup> de pacientes con LMC al diagnóstico vs. CD34<sup>+</sup> de dadores sanos (de leucaféresis o cordón umbilical)<sup>95</sup>.

## V.2. Análisis de perfiles de expresión

Con los listados de microARNs surgidos de la secuenciación masiva se realizó un análisis de las posibles vías de señalización diferencialmente involucradas en cada población. Para ello se utilizó la herramienta miRPath (*DIANA tools*), la cual realiza una búsqueda de los blancos de cada microARN (mediante métodos de predicción o búsqueda de interacciones experimentalmente validadas), asocia dichos blancos a las vías de señalización KEGG (*Kyoto Encyclopedia of Genes and Genomes*) y términos GO (*Gene ontology*), y finalmente realiza un análisis de agrupamiento de microARNs (*clustering*) y enriquecimiento significativo de vías KEGG o términos GO. Para este análisis sólo se tuvieron en cuenta las interacciones microARN-*target* experimentalmente validadas (almacenadas en Tarbase), exceptuando aquellos microARNs para los cuales no había registro de tal tipo; en esos casos se seleccionó la herramienta de predicción de blancos microT-CDS (ver Materiales y métodos).

En el caso de los microARNs diferencialmente expresados en la fracción de LSC de pacientes (pregunta A, capítulo IV), se observaron tres agrupamientos de microARNs (Fig. 5.1):

*cluster 1:* miR-10a-5p y miR-125a-5p: enriquecido principalmente en vías de "metabolismo de ácidos grasos", "biosíntesis de ácidos grasos", "vía Hippo";

*cluster 2:* miR-708-5p, miR-2355-5p, miR-99b-5p, miR-196a-5p, miR-92b-3p, miR-126-5p: enriquecido principalmente en vías de "interacción receptor-matriz extracelular", "uniones adherentes", "proteoglicanos en cáncer", vía "Hippo";

*cluster 3:* miR-485-5p, miR-382-5p, miR-411-5p, miR-432-5p, miR-409-3p, miR-323b-3p, miR-431-5p, miR-134-5p: enriquecido en vías de "interacción receptor-matriz extracelular", "metabolismo de xenobióticos por citocromo P450", "enfermedades causadas por priones".

Los agrupamientos 1 y 2 contienen microARNs sobreexpresados en LSC, con excepción de miR-708-5p. El agrupamiento 3 contiene microARNs subexpresados en LSC, con excepción de miR-411-5p (Fig. 5.1).

Las vías enriquecidas significativamente se detallan en la tabla 5.1.

Tabla 5.1. Vías de señalización asociadas a los blancos de los microARNs diferencialmente expresados en LSC de LMC. Se muestra el número de genes y microARNs asociados a cada vía. Se consideraron sólo aquellos resultados con  $p$ -valor < 0,001.

Vía KEGG	p-valor	#genes	#microARNs
Enfermedades causadas por priones	0	1	1
Biosíntesis de ácidos grasos	0	1	2
Interacción receptor-matriz extracelular	0	15	5
Metabolismo de xenobióticos por citocromo P450	1,42E-10	3	1
Proteoglicanos en cáncer	1,84E-10	58	4
Vía de señalización Hippo	4,5E-06	39	2
Metabolismo de ácidos grasos	9,12E-06	4	2
Uniones adherentes	5,66E-05	17	1

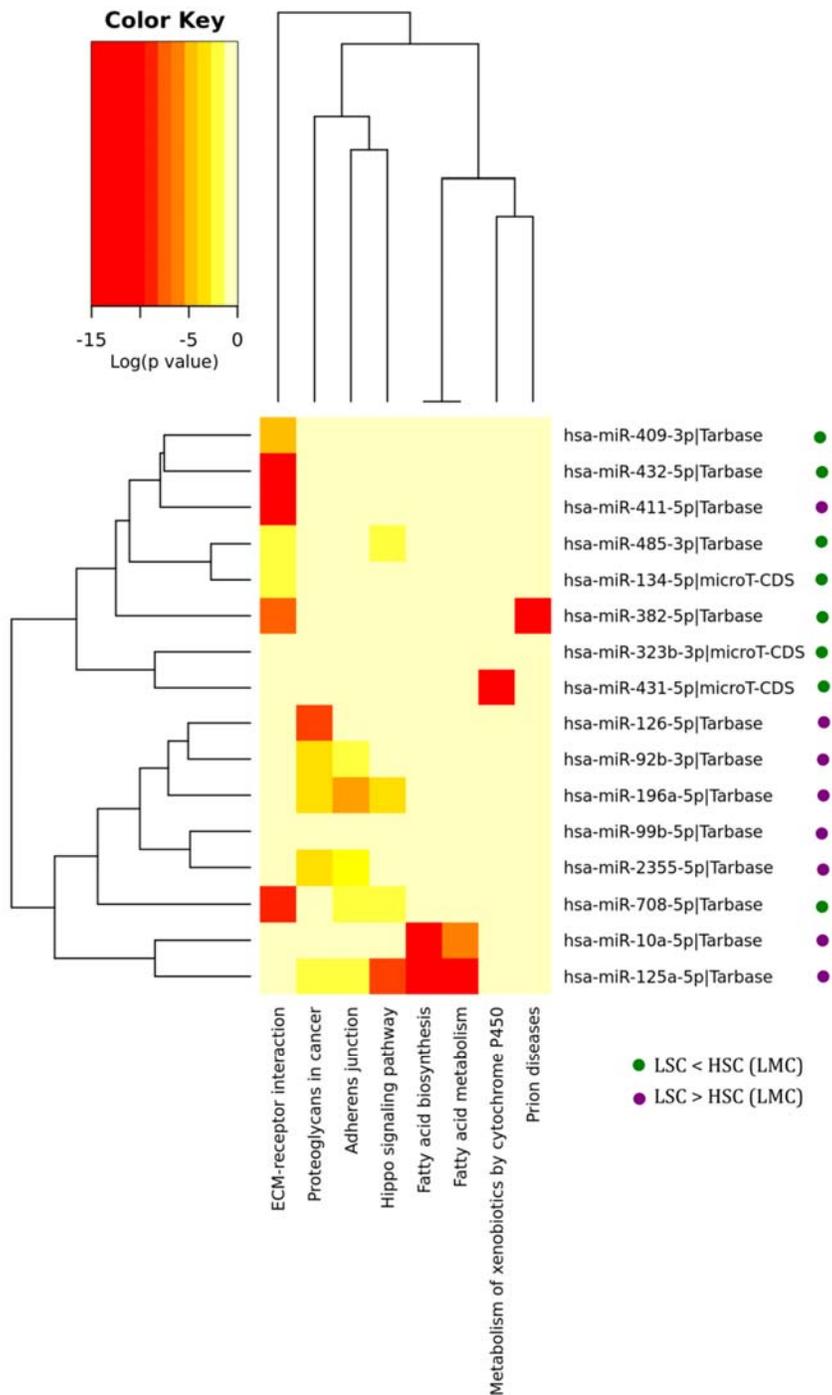


Fig. 5.1. Mapa de color de vías de señalización enriquecidas en el listado de microARNs diferencialmente expresados en LSC de LMC (miRPath, pathway union).

En el caso de los microARNs diferencialmente expresados en la fracción de HSC de pacientes (pregunta B, capítulo IV, listado reducido), se observaron dos agrupamientos de microARNs (Fig. 5.2):

*cluster 1:* miR-224-5p, miR-181c-5p, miR-183-5p, miR-182-5p, miR-195-5p, miR-20a-5p, let-7f-5p, let-7a-5p. Se observó un enriquecimiento en vías asociadas a distintos

procesos proliferativos como "carcinogénesis viral", "vías en cáncer", "leucemia mieloide crónica". Los resultados también incluyeron vías enriquecidas en las LSC de pacientes, como son "biosíntesis y metabolismo de ácidos grasos", "uniones adherentes", "vía Hippo", "proteoglicanos en cáncer", e "interacción receptor-matriz extracelular", sugiriendo características comunes con dicha fracción.

*cluster 2:* miR-1304-5p, miR-629-3p, miR-362-5p, miR-548o-3p, miR-629-5p, miR-144-3p, miR-654-3p, let-7c-5p. Se observó un enriquecimiento principalmente en la vía de "biosíntesis de esteroides".

El agrupamiento 1 contiene microARNs subexpresados en la fracción de HSC de pacientes, mientras que el cluster 2 contiene microARNs con ambas tendencias (Fig. 5.2). El listado de las vías enriquecidas significativamente se detalla en la tabla 5.2.

*Tabla 5.2. Vías de señalización asociadas a los blancos de los microARNs diferencialmente expresados en HSC de LMC. Se muestra el número de genes y microARNs asociados a cada vía. Se consideraron sólo aquellos resultados con p-valor < 0,001.*

Vía KEGG	p-valor	#genes	#microARNs
<b>Biosíntesis de ácidos grasos</b>	0	3	2
<b>Vía señalización Hippo</b>	0	62	5
<b>Carcinogénesis viral</b>	0	98	5
<b>Interacción receptor-matriz extracelular</b>	8,88E-10	20	3
<b>Metabolismo de ácidos grasos</b>	1,22E-09	8	2
<b>Degradación de lisina</b>	1,22E-09	23	4
<b>Uniones adherentes</b>	2,25E-06	42	4
<b>Hepatitis B</b>	5,09E-06	73	5
<b>Ciclo celular</b>	3,33E-05	68	4
<b>Proteoglicanos en cáncer</b>	1,98E-03	80	4
<b>Meiosis en oocitos</b>	3,87E+0	47	2
<b>Vías en cáncer</b>	0	4,76E+0	67
<b>Glioma</b>	0	1,09E+0	23
<b>Biosíntesis de esteroides</b>	1	6,96E+0	3
<b>Procesamiento proteínas en retículo endoplásmico</b>	1	9,12E+0	40
<b>Leucemia mieloide crónica</b>	1	0,00016	34
	78296		2

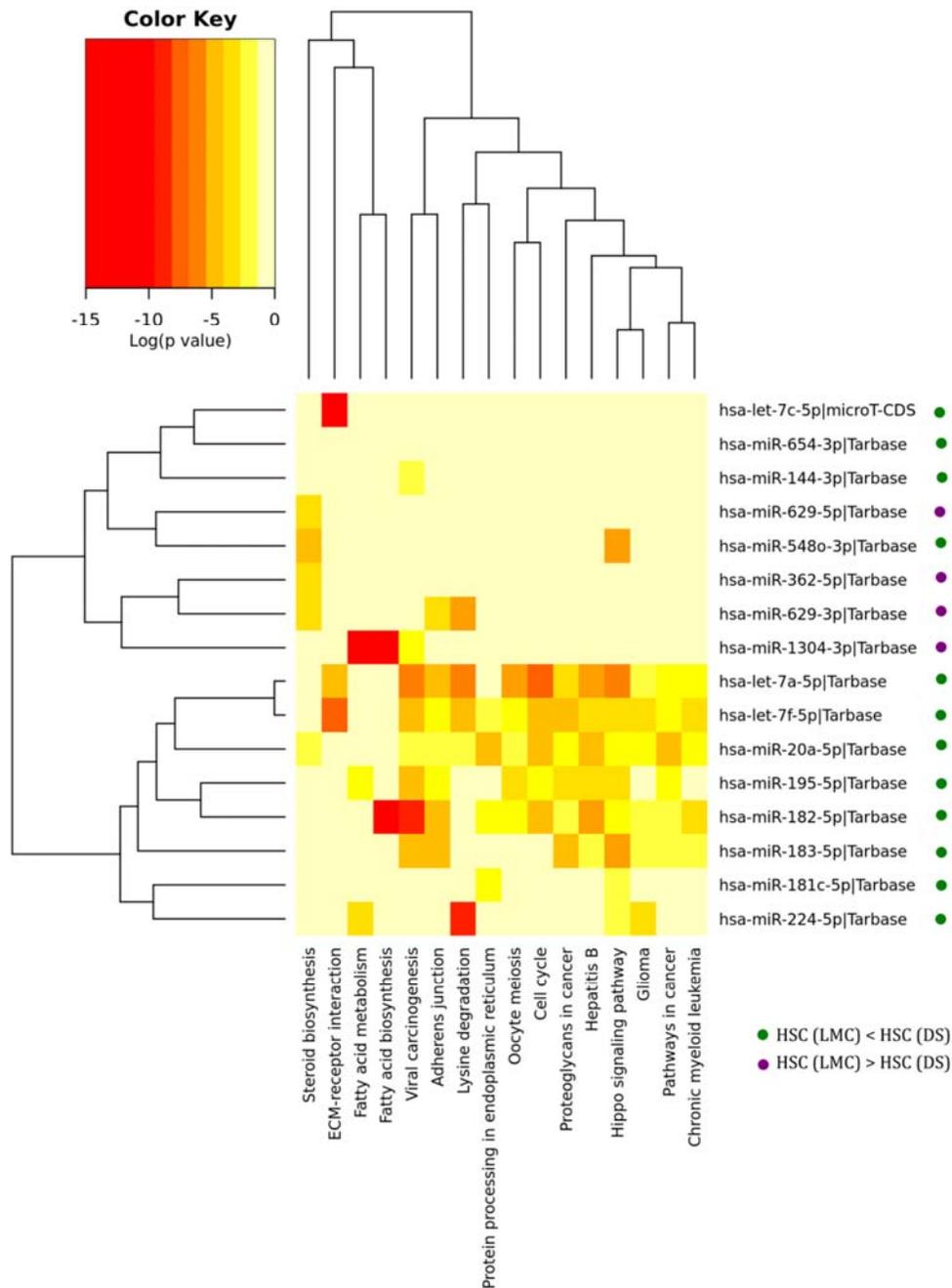
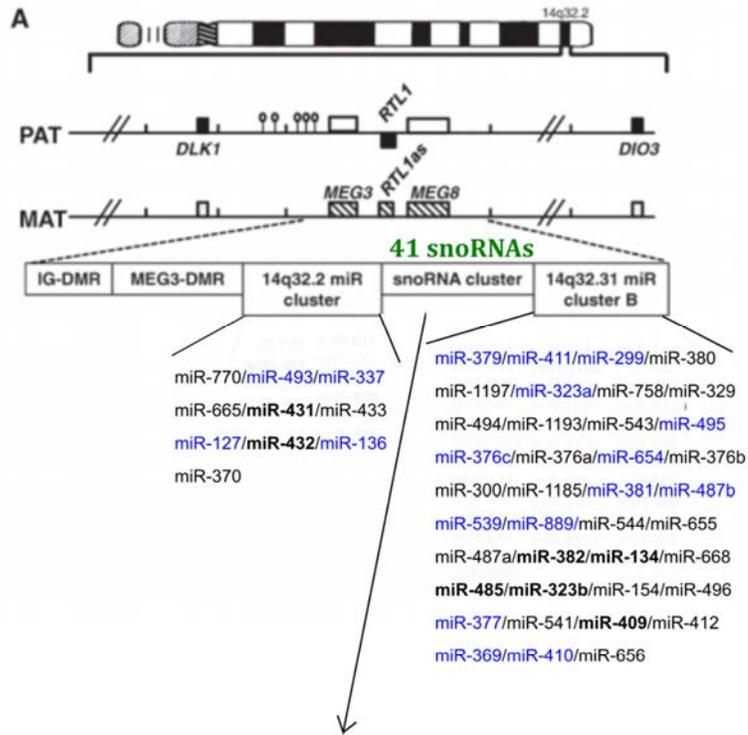


Fig. 5.2. Mapa de color de vías de señalización para el listado de microARNs diferencialmente expresados en HSC de LMC (miRPath, pathway union)

Estos resultados sugieren que las HSC de pacientes presentan una desregulación a nivel de varias vías de señalización, diferenciándose de las HSC de DS. Llamativamente, fue mucho mayor el número de microARNs y vías desreguladas en esta comparación que en la pregunta A), sugiriendo una mayor divergencia fenotípica. La observación de que esta población presenta una desregulación del miRnoma implicaría que se trata de células con un fenotipo alterado, posiblemente afectado por la coexistencia con las LSC y otras células leucémicas.

Por otro lado, los microARNs subexpresados en LSC resultaron, en general, ser los de menor abundancia relativa, y similar entre sí (Capítulo IV, Tabla 4.3). Analizando la localización genómica de éstos, encontramos que todos, excepto miR-708-5p, pertenecen a los *clusters* genómicos miR-379/miR-410 y miR-127/miR-136, ambos localizados en una región del cromosoma 14 (14q32) caracterizada por sufrir *imprinting* (locus Dkl1/DIO3)<sup>125</sup>. La observación de que los conteos de estos microARNs son menores en la fracción de LSC en comparación con todas las restantes sugiere que esta región podría estar silenciada en las LSC por algún mecanismo epigenético. De acuerdo con esta observación, analizando los ARNs pequeños nucleolares (snoARNs) presentes en dicho locus, se observó que 5/41 se encontraban subexpresados ( $GFOLD \geq |1|$ ) en LSC o HSC de pacientes, en comparación con HSC de donantes sanos (Fig. 5.3). Revisando el GFOLD de todos los microARNs de ambos *clusters*, se detectó que un gran número de ellos se encontraba subexpresado ( $GFOLD \geq |1|$ ) en LSC vs. HSC de DS, aunque no pertenecían al listado obtenido en la pregunta A debido a que los niveles de éstos no variaban entre las fracciones LSC y HSC de pacientes (Fig. 5.3).



ARN pequeño nucleolar	GFOLD (HSC DS vs. LSC)	GFOLD (HSC DS vs. HSC LMC)
14q(II-24)	0,6	1,1
14q(II-21)	1,0	0
14q(II-15)	1,1	1
14q(II-22)	3,4	2,5
14q(II-9)	3,5	2,2

Fig. 5.3. Representación gráfica de la región genómica que contiene a los clusters de microARNs 14q32.2 y 14q32.1, y a un cluster de 41 snoARNs. Los genes con imprinting paterno (DLK1, RTL1, DIO3) están en negro, y con imprinting materno (MEG3, anti-RTL1, MEG8) en líneas diagonales. Se indican en negrita aquellos microARNs subexpresados en LSC vs. HSC de DS y vs. HSC de pacientes, en azul aquellos subexpresados en LSC vs. HSC de DS, pero sin diferencias con respecto a las HSC de pacientes. Del total de snoARNs presentes en el cluster, 5 se encontraron subexpresados (GFOLD  $\geq$ 1) en LSC o HSC de pacientes, en comparación con HSC de DS. Modificado de<sup>126</sup>.

Por último, dado que las asociaciones realizadas fueron basadas en aquellas interacciones experimentalmente validadas (Tarbase), se consideró la posibilidad de que los resultados obtenidos estén sesgados. Para poner a prueba este punto, se sometió al mismo análisis *in silico* a dos listados de 16 microARNs seleccionados al azar con GFOLD=0 (pero detectables en ambas fracciones) en la comparación LSC vs. HSC de pacientes. Se obtuvo un menor número de vías, microARNs y genes para ambos listados (tabla 5.3), dando soporte a los resultados obtenidos previamente.

Tabla 5.3. Vías de señalización asociadas a los targets de dos listados de microARNs con GFOLD=0 entre LSC y HSC de LMC. Se consideraron sólo aquellos resultados con p-valor<0,001.

Vía KEGG	p-valor	#genes	#microARNs
<b>Biosíntesis de ácidos grasos</b>	7,77E-16	1	1
Listado 1: hsa-mir-3136, hsa-mir-543, hsa-mir-3202, hsa-mir-1256, hsa-mir-1976, hsa-mir-1185-1-3p, hsa-mir-639, hsa-mir-550a-3p, hsa-mir-548b-5p, hsa-mir-3136-5p, hsa-mir-939-5p, hsa-mir-3134 hsa-mir-375, hsa-mir-1908-5p, hsa-mir-202-5p, hsa-mir-885-5p			
Vía KEGG	p-valor	#genes	#microARNs
<b>Interacción extracelular</b>	<b>receptor-matriz</b> 0	22	3
<b>Vía señalización Hippo</b>	2,21E-01	7	2

### V.3. Análisis del novel-3

Sorpresivamente, el alineamiento de esta secuencia con el genoma humano (BLASTN) arrojó como resultado que el mismo se encuentra codificado en el cromosoma 15, en la región y cadena correspondiente a SCARNA15 (*small Cajal body-specific RNA 15*). Se trata de un ARN de 127pb no codificante perteneciente a la categoría de ARNs pequeños nucleolares (snoARNs). A su vez, una búsqueda en miRBase reveló homología de secuencia con miR-1839 presente en otras especies, mayoritariamente de mamíferos (Fig. 5.4).

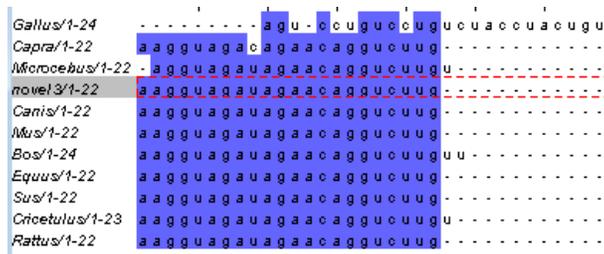


Fig. 5.4. Alineamiento múltiple de secuencias (CLUSTAL OMEGA) entre la secuencia del novel-3 y las correspondientes a miR-1839 en otras especies reportadas en miRBase. En azul se muestran los nucleótidos con mayor conservación en las especies comparadas.

Con el objetivo de predecir posibles funciones, se realizó una búsqueda en miRPath de vías asociadas a miR-1839 en la especie *Mus musculus* (ratón) (mmu-miR-1839-5p). Se encontró una interacción experimentalmente validada entre miR-1839 y un gen asociado a "vías de señalización que regulan la multipotencia de células madre" (Skil) (tabla 5.4), sugiriendo que la molécula descrita podría ser de interés para el subtipo de células en estudio. Por otro lado, en el genoma de ratón, este microARN también coincide con un scaRNA denominado "Mir1839-201 scaRNA".

Tabla 5.4. Vías de señalización asociadas a los blancos experimentalmente validados de mmu-miR-1839-5p. Se consideraron sólo aquellos resultados con p-valor<0,001.

Vía KEGG	p-valor	#genes	#microARNs
Reabsorción de agua regulada por vasopresina	0,0124	1	1
<b>Vías de señalización que regulan la multipotencia de células madre</b>	0,0188	1	1
Enfermedad de Huntington	0,0188	1	1

#### V.4 Análisis de vías asociadas a los microARNs validados por RT-qPCR.

Con el objetivo de diseñar ensayos funcionales que nos permitan evaluar el efecto en el fenotipo ocasionado por la desregulación de los microARNs validados, se realizó una búsqueda extensa en bibliografía y mediante métodos bioinformáticos de posibles funciones biológicas. En esta búsqueda se incluyeron aquellos microARNs cuya prueba global fue estadísticamente significativa: miR-196a-5p, miR-125a-5p, miR-92b-3p y miR-10a-5p. Se excluyó a miR-126-5p por encontrarse extensamente estudiado en LSC de AML y LMC <sup>124,127</sup>, aunque se discutirá brevemente su función en la discusión de este capítulo.

Los análisis consistieron en:

- 1) Predicción de los posibles ARNm blancos de los 4 microARNs en simultáneo mediante 10 herramientas bioinformáticas de predicción (DIANA-microT, miRanda, miRDB, RNAhybrid, pictar(4way), pictar(5way), PITA, RNA22, TargetScan, miRWalk) con la herramienta ChmiRs<sup>73</sup>;
- 2) Búsqueda de vías moleculares (KEGG) asociadas a los ARNm blancos predichos (ChmiRs y miRPath)
- 3) Búsqueda de términos biológicos (GO) asociados a los ARNm blancos predichos (ChmiRs y miRPath)
- 4) Búsqueda en mirTarbase de los microARNs individuales. Esta base de datos almacena evidencias experimentales de asociaciones microARN-ARNm por distintos ensayos funcionales<sup>128</sup>.

Los resultados más relevantes se listan en la Tabla 5.5, ordenados según la significancia estadística calculada por el algoritmo de búsqueda.

Tabla 5.5. Ranking de las principales vías KEGG y términos GO asociados a hsa-miR-10a-5p, hsa-miR-125a-5p, hsa-miR-92b-3p y hsa-miR-196a-5p. Se muestran resaltados los términos asociados a algún tipo de proceso metabólico.

Vías KEGG (ChemiR)	Vías KEGG (miRPath-Tarbase)	Términos GO (ChemiR)	Categorías GO (miRPath-microT-CDS)
1. <b>Metabolismo</b>	A. <b>Biosíntesis y metabolismo de ácidos grasos</b>	1. <b>Glucólisis canónica</b>	1. <b>Organela</b>
2. <b>Cáncer vejiga</b>	B. Vía de señalización Hippo, LMC, uniones adherentes	2. <b>Homeostasis de la glucosa</b>	2. <b>Procesos celulares metabólicos de compuestos nitrogenados</b>
3. <b>Detención y recuperación de la elongación (transcripción)</b>	C. Degradación de lisina, proteoglicanos en cáncer	3. <b>Gránulos ribonucleoproteicos citoplasmáticos</b>	3. <b>Procesos biosintéticos</b>
4. <b>Metabolismo de fructosa y manosa</b>	D. <b>Biosíntesis de valina, leucina e isoleucina; biosíntesis de otros tipos de O-glicanos</b>	4. Regulación negativa de la transición epitelio-mesenquimal	
5. <b>Biosíntesis de ácidos grasos insaturados</b>	E. Vías de señalización que regulan la multipotencia de células madre	5. Regulación de la transcripción involucrada en la transición G1/S del ciclo mitótico	
6. <b>Metabolismo de galactosa</b>		6. Unión de factores de crecimiento	
7. <b>Ritmo circadiano - mamíferos</b>		7. Unión de GTPasa Ran	
8. <b>Metabolismo de ácido alfa-linoleico</b>		8. Morfogénesis del sistema esquelético	
9. <b>Translocación de SLC2A4(Glut4) a la membrana plasmática</b>		9. Proceso metabólico de ARN no codificante	
10. <b>Exportación de proteínas (al exterior celular)</b>		10. Liberación de citocromo c de la mitocondria	

Resultó llamativo el enriquecimiento en vías relacionadas con distintos tipos de metabolismo (vías KEGG-ChemiR: 5/10, vías KEGG-miRPath-Tarbase: 2/5, términos GO-ChemiR: 2/10, categorías GO-miRPath-microT-CDS: 2/3), y en particular, procesos anabólicos o de biosíntesis. Suponiendo el mecanismo canónico de acción de los microARNs, ésto implicaría una disminución en los niveles de los ARNm blanco y/o de las proteínas codificadas por ellos; en este caso resultaría en una inhibición o atenuación de procesos biosintéticos. En línea con este hallazgo, recientemente se ha reportado que la fracción enriquecida en células primitivas/progenitoras de pacientes con LMC depende de un metabolismo oxidativo aumentado: la comparación del metaboloma entre las fracciones CD34<sup>+</sup> y CD34<sup>-</sup> de pacientes (y también vs. CD34<sup>+</sup> de donantes sanos)

evidenció un aumento en glicerol-3-fosfato y derivados de carnitina y acilcarnitina, y una disminución de ácidos grasos libres. Los autores realizaron también ensayos funcionales, en los que demostraron un aumento en la tasa de consumo de oxígeno mitocondrial, y un aumento de metabolitos del ciclo de Krebs y sus aminoácidos derivados (ej. glutamato y aspartato)<sup>129</sup>.

Este resultado resalta el efecto pleiotrópico de los microARNs y su potencial como biomarcadores: la búsqueda bioinformática de blancos predichos para sólo 4 microARNs sugirió un fenotipo alterado que concuerda con datos experimentales provenientes de ensayos funcionales.

#### **V.5. Selección de potenciales genes blanco de los microARNs validados por RT-qPCR.**

Con el objetivo de investigar genes blanco candidatos de los microARNs desregulados, se realizó una selección de aquellos relacionados con procesos metabólicos. Se utilizaron tres listados de genes, pertenecientes a las siguientes listas *hallmark*<sup>130</sup>:

**AG: "hallmark metabolismo de ácidos grasos"**

**FO: "hallmark fosforilación oxidativa"**

**GL: "hallmark glucólisis"**

Se realizaron intersecciones entre los siguientes conjuntos (Fig. 5.5):

- 1: potenciales ARNm blanco (predicción por 10 métodos, ChemiRs)
- 2: genes pertenecientes a cada categoría *hallmark* (3 listas)
- 3: ARNm desregulados <sup>96</sup>
- 4: proteínas desreguladas <sup>95</sup>

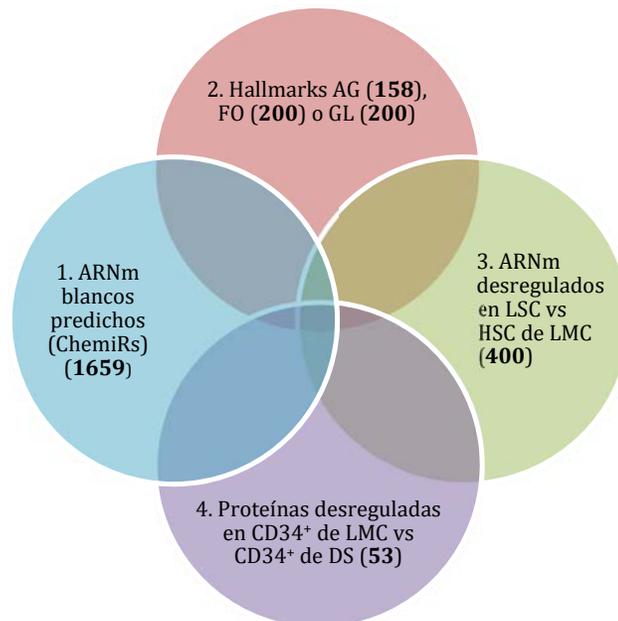


Fig. 5.5. Esquema de los conjuntos analizados indicando el número de genes en cada grupo.

Se obtuvieron los siguientes resultados:

**AG: hallmark "metabolismo de ácidos grasos"**

1) Un único gen en común entre las cuatro listas: **UROD** (uroporphyrinogen decarboxylase), enzima involucrada en la síntesis del hemo. A nivel de ARNm y proteína, este gen se encontró sobreexpresado en la fracción de LSC (ARNm) y CD34<sup>+</sup> LMC (proteína). El resultado de ChemiRs predice la interacción entre UROD y hsa-miR-92b-3p.

2) Doce genes en común entre las listas 1 y 2. AG: UGDH, RDH11, ACAA1, FASN, SLC22A5, HSP90AA1, GLUL, GRHPR, ALDOA, UROD, DLST, PTPRG. Tres de ellos son predichos como genes blanco por varios métodos: **UGDH** (UDP-glucose 6-dehydrogenase; 5 métodos para hsa-miR-10a-5p), **RDH11** (Retinol dehydrogenase 11; 5 métodos para hsa-miR-196a-5p) y **ACAA1** (Acetyl-CoA Acyltransferase 1; 4 métodos para hsa-miR-92b-3p).

**GL: hallmark "glucólisis"**

3) No se encontraron genes en común entre las cuatro listas. Se encontró un único gen en común entre la lista 2. GL, los potenciales ARNm blanco de los microARNs, y los ARNm desregulados: **PGAM1** (phosphoglycerate mutase 1). A nivel de ARNm, este gen se encontró sobreexpresado en la fracción de LSC. El resultado de ChemiRs predice interacción con hsa-miR-92b-3p y hsa-miR-196a-5p.

4) 18 genes en común entre las listas 2. GL y 1: ALDOA, ENO1, TPI1, VEGFA, HK2, PGAM1, PC, G6PD, EGFR, GOT2, SLC25A13, PDK3, B4GALT1, DDIT4, AURKA, PAM, PYGL, FKBP4. Cinco de ellos son predichos como genes blanco por varios métodos: **VEGFA** (vascular endothelial growth factor), **HK2** (hexokinase-2) y **B4GALT1** (beta-1,4-galactosyltransferase 1), fueron predichos por 5 métodos para hsa-miR-125a-5p; **DDIT4** (DNA damage inducible transcript 4) y **AURKA** (aurora kinase A) fueron predichos por 4 métodos para hsa-miR-92b-3p.

**FO: hallmark "fosforilación oxidativa"**

5) No se encontraron genes en común entre las cuatro listas. Se encontraron diez genes en común entre las listas 1, 3 y 2. FO: ATP5G3, COX7B, NDUFB6, NDUFB5, ATP5F1, ATP5J2, VDAC3, VDAC1, PHB2, SLC25A5. **NDUFB6** (NADH:Ubiquinone Oxidoreductase Subunit B6) y **ATP5F1** (ATP Synthase Peripheral Stalk-Membrane Subunit B) son predichos por 5 métodos para hsa-miR-10a-5p. Todos ellos se detectaron sobreexpresados a nivel de ARNm en la fracción de LSC.

6) 24 genes en común entre las listas 2. FO y 1: ATP5G3, COX7B, NDUFB6, UQCRC2, NDUFB5, ATP6V1F, ATP5F1, ATP5J2, VDAC3, ATP5G2, VDAC2, ABCB7, VDAC1, PHB2, GOT2, SLC25A5, DLST, SLC25A4, ACAA1, TIMM50, MRPL35, GPI, OAT. Ocho de ellos son predichos como genes blanco por varios métodos: **NDUFB6** (NADH:Ubiquinone Oxidoreductase Subunit B6), **ATP5F1** (ATP Synthase Peripheral Stalk-Membrane Subunit B) y **TIMM50** (Translocase Of Inner Mitochondrial Membrane 50), fueron predichos por 5 métodos para hsa-miR-10a-5p; **ATP6V1F** (ATPase H<sup>+</sup> Transporting V1 Subunit F) y **MRPL35** (Mitochondrial Ribosomal Protein L35) fueron predichos por 4 métodos para hsa-miR-196a-5; **ATP5G2** (ATP Synthase Membrane Subunit C Locus 2) fue predicho por 6 métodos para hsa-miR-125a-5p y **ACAA1** (Acetyl-CoA Acyltransferase 1) fue predicho por 4 métodos para hsa-miR-92b-3p.

7) Independientemente de las listas *hallmark*, se encontraron 3 genes en común adicionales entre las proteínas desreguladas y los potenciales ARNm blanco de los microARNs: **ABR**, **BAK1**, y **CPD**. Sólo ABR se detectó subexpresado a nivel proteico en la fracción leucémica. En particular, ChemiR predice la interacción entre **ABR (Active breakpoint cluster region-related protein)** y **hsa-miR-92b-3p** por 5 métodos. En mirTarbase se reportan evidencias experimentales de la interacción entre hsa-miR-92b-3p, la región 3'UTR de ABR y AGO1 en la línea celular HEK293<sup>131</sup>. También se observó una correlación negativa entre los niveles de microARN y ARNm en un set de datos correspondiente a células madre embrionarias diferenciadas (R=-0,555; p=0,003; GSE42095:" regulación dinámica de las firmas de microARN y ARNm durante la diferenciación pancreática *in vitro*"), sugiriendo una posible relevancia funcional de esta interacción.

## V. 6. Conclusiones parciales y comentarios

A continuación, se discuten las principales conclusiones extraídas a partir de los resultados de esta tercera parte:

1) *Las LSC y HSC presentes en pacientes con LMC al diagnóstico presentan un patrón de disminución global de los niveles de microARNs con respecto a las HSC de DS.*

Por secuenciación masiva se observó que la mayoría de los microARNs expresados diferencialmente (GFOLD $\geq$ |2|) se encontraban subexpresados en pacientes vs. donantes sanos: 88% en la comparación LSC vs. HSC de DS (Fig. 5.6, a), y 86% en la comparación HSC de LMC vs. HSC de DS (Fig. 5.6, b). Este patrón concuerda con lo reportado en bibliografía como disminución global de los niveles de microARNs en cáncer, independientemente del tejido, tanto en muestras de pacientes como líneas celulares<sup>132</sup>. En cambio, en la comparación de LSC vs. HSC de pacientes (Fig. 5.6, c), el 23% de los microARNs están subexpresados en la fracción de LSC. Estas tendencias se mantienen si se consideran los microARNs con GFOLD $\geq$ |1|. Si se comparan las fracciones correspondientes a DS, el 67% de los microARNs se subexpresan en los progenitores (Fig. 5.6, d).

a. LSC vs HSC DS	b. HSC LMC vs HSC DS	c. LSC vs HSC LMC	d. Prog. vs HSC DS
2,876	-3,63657	-4,11205	-2,81167
-2,70695	-3,47663	-4,00851	-2,80133
-2,0313	-3,08492	-3,62383	-2,57077
2,01793	-3,0391	-3,61191	-2,40533
2,07772	-2,36923	-3,61076	-2,33622
2,1318	-2,06799	-3,5973	-2,04696
2,15527	2,03388	-3,58063	-2,015
<b>2,27472</b>	<b>2,04624</b>	<b>-3,54201</b>	<b>-2,00936</b>
2,28768	2,07654	-3,40483	-2,00071
2,34628	2,11522	-3,37445	2,01016
2,37414	2,14404	-3,32667	2,03954
2,45461	2,20212	-3,32302	2,04612
2,47879	2,27388	-3,27343	2,13256
2,67221	2,31173	-2,90772	2,40396
2,91919	2,32827	-2,90588	2,45806
2,98316	2,34854	-2,84068	2,47053
3,11675	2,35273	-2,74264	2,56067
3,35487	2,35319	-2,72374	2,56658
3,40507	2,4829	-2,7137	2,56855
3,41177	2,4957	-2,70218	2,61266
3,59807	2,53006	-2,59439	2,63659
3,63084	2,58483	-2,58398	2,64317
3,71306	2,61444	-2,5776	2,9829
3,75513	2,66894	-2,47304	3,36366
3,76322	2,71658	-2,4439	3,67
4,3887	2,74879	-2,34491	3,97734
	2,75793	-2,33048	4,08676
	2,90637	-2,30851	4,78075
	2,93952	-2,30791	
	3,0569	-2,27605	
	3,08552	-2,1718	
	3,08671	-2,1378	
	3,08789	-2,12024	
	3,10694	-2,11358	
	3,20363	-2,10729	
	3,33779	-2,10408	
	3,50008	-2,08634	
	3,63681	-2,07361	
	3,66252	-2,04242	
	3,79038	-2,01677	
	3,85427	-2,01672	
	3,9226	2,08765	
	5,13848	2,14498	
	5,44698	2,29828	
		2,53122	
		2,64894	
		2,80722	
		2,89645	
		2,91926	
		2,96473	
		3,39316	
		3,66335	
		3,82733	

Fig. 5.6. Mapa de color de los valores de GFOLD para las cuatro comparaciones realizadas con los datos provenientes de la secuenciación masiva. Se consideraron sólo aquellos valores de  $GFOLD \geq |2|$  para simplificar su representación gráfica. Las filas NO se corresponden con los mismos microARNs. Los valores de  $GFOLD < 0$  (color verde) implican mayores niveles en la fracción de LSC (comparaciones a y c), HSC de LMC (comparación b), o progenitores de DS (comparación d). Los valores de  $GFOLD > 0$  (color amarillo-naranja) implican menores niveles en la fracción de LSC (comparaciones a, y c), HSC de LMC (comparación b) o progenitores de DS (comparación d).

Respecto a la disminución global de los niveles de microARNs en cáncer, aún se desconocen los mecanismos subyacentes. Un trabajo reciente ha demostrado, en líneas celulares, que durante la pérdida de respuesta a la confluencia (asociada al proceso de transformación), la desregulación en la vía Hippo ocasiona la presencia de la proteína YAP en el núcleo, capaz de secuestrar a p72, proteína involucrada en la biogénesis de microARNs. Este mismo mecanismo, en condiciones fisiológicas, también estaría involucrado en la disminución de los niveles globales de microARNs en células en proliferación, con la vía Hippo funcionando correctamente<sup>133</sup>. Este mecanismo podría explicar el patrón observado en las comparaciones de fracciones de pacientes con DS (Fig. 5.5, a-b), y entre fracciones de DS (Fig. 5.5, d), ya que en este último caso, los progenitores son más proliferativos. Sin embargo, el patrón observado en la comparación de fracciones LSC y HSC, ambas de pacientes, parece contradecir el modelo y resulta difícil de interpretar, sugiriendo la existencia de mecanismos alternativos.

2) El análisis *in silico* de los perfiles de expresión de microARNs obtenidos por secuenciación masiva reveló un enriquecimiento para aquellos microARNs subexpresados en las HSC de pacientes, en vías asociadas a procesos proliferativos.

Este resultado sugiere que las HSC presentes en pacientes, a pesar de no albergar la mutación *BCR-ABL1*, presentarían un fenotipo alterado en comparación con aquellas HSC provenientes de DS, sugiriendo la existencia de factores extrínsecos, probablemente asociados a la coexistencia con células leucémicas. Por ejemplo, se ha postulado que las células mesenquimales de la MO son capaces de transferir microARNs en vesículas extracelulares a células del sistema hematopoyético<sup>124</sup>. El análisis de enriquecimiento en vías moleculares resultó en una gran cantidad de vías proliferativas asociadas a los microARNs subexpresados en las HSC de pacientes, lo que sugiere una desregulación de la proliferación de las mismas.

Resultó llamativo también que el número total de microARNs desregulados fue mayor en esta comparación que en la de LSC vs. HSC de DS (Figs. 4.11 y 4.12). Una posible explicación a este patrón es la inclusión de la población  $CD38^{dim}$  en la fracción HSC de DS, la cual incluye a progenitores multipotentes. Un estudio reciente caracterizó células únicas provenientes de muestras de MO de pacientes con LMC al diagnóstico (n=1263 células en total) por RT-qPCR, lo que permitió la identificación de 7 subpoblaciones en la fracción  $Lin-CD34^+CD38^{-/low}$  por análisis de agrupamiento no supervisado, que reflejaban su estado de diferenciación y proliferación (subpoblaciones mieloides I a IV, linfoide, megacariocito/eritroide y primitivas). Mientras que la población  $BCR-ABL^+$  estaba conformada mayoritariamente por células de las subpoblaciones mieloides III/IV (estadios más diferenciados), la población  $BCR-ABL^-$  presentó una predominancia de células de las subpoblaciones primitiva y mieloides I<sup>97</sup>. Esta correlación entre diferenciación y *BCR-ABL1* concuerda con observaciones previas, tanto nuestras (capítulo III, Fig. 3.1) como de otros autores<sup>88</sup>. Por citometría de flujo, hemos detectado también que las LSC presentaban un patrón de  $CD45$  vs. SSC solapado con el de las células progenitoras  $CD38^+$  (Fig. 5.7).

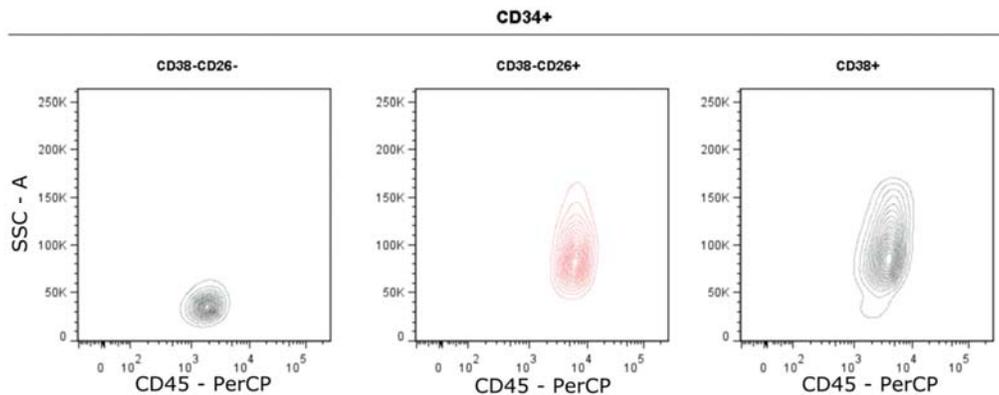


Fig. 5.7. Perfil representativo de los niveles del marcador de superficie  $CD45$  vs. complejidad interna (SSC) en un paciente con LMC al diagnóstico. Se comparan las subpoblaciones  $CD38-CD26^-$  (HSC) y  $CD38-CD26^+$  (LSC), con la población de progenitores  $CD38^+$ .

3) La mayoría de los microARNs subexpresados en las LSC en comparación con las HSC de pacientes, pertenecen al locus *Dlk1-DIO3*, el cual puede sufrir imprinting.

Este resultado podría ser de potencial relevancia para el descubrimiento de mecanismos novedosos de regulación de la expresión génica en las LSC, ya que existen reportes de

que la expresión de dicho locus está enriquecida en la fracción de LT-HSCs, y que los microARNs codificados en dicho locus suprimen componentes de la vía PI3K-mTOR, restringiendo el metabolismo mitocondrial, mientras que el silenciamiento de este locus por *imprinting* conlleva defectos funcionales en HSC de hígado fetal<sup>134</sup>. La observación de que los microARNs de este locus están subexpresados en las LSC, de ser confirmada en otra cohorte de pacientes, sugeriría que existen mecanismos epigenéticos de silenciamiento en las LSC, y que las consecuencias funcionales de dicho silenciamiento podrían estar relacionadas con algunas de las vías desreguladas en LSC, como la vía PI3K-Akt<sup>135</sup> y el metabolismo oxidativo<sup>129</sup>.

*4) El novel-3 detectado por secuenciación masiva y RT-qPCR corresponde a un ARN pequeño derivado de SCARNA15, y tiene homología de secuencia con miR-1839 de otras especies.*

Los scaRNAs (*small Cajal body RNA*) son un subtipo de ARN pequeños nucleolares (snoARNs). Los snoARNs (60-300nt) han sido tradicionalmente implicados en procesos de maduración de otras especies de ARN (principalmente modificación post-transcripcional de ARNs ribosomales y algunos ARNs que forman parte del spliceosoma), a través de su asociación con proteínas para formar partículas ribonucleoproteicas de localización nucleolar. Los scaRNAs se localizan en los cuerpos de Cajal, organelas subnucleares localizadas en el nucleoplasma. En los últimos años se han comenzado a estudiar los snoARNs/scaRNAs como moléculas con posible asociación a procesos patológicos, aunque su relevancia y los mecanismos implicados aún no han sido descritos. Se ha observado una disminución global de estas especies en muestras de pacientes con mieloma múltiple<sup>136</sup> y patrones específicos de expresión en leucemia mieloide aguda<sup>137</sup>.

En el caso del ARN pequeño derivado de SCARNA15 (también denominado ACA45), se ha demostrado que puede ser procesado por un mecanismo independiente de DGCR8/Drosha y dependiente de Dicer, puede unir Ago1 y Ago2, y actuar como regulador post-transcripcional de CDC2L6. En el trabajo lo denominan ACA45 sRNA (*ACA45 small RNA*)<sup>138</sup>. En ratón, miR-1839 también está incluido en un scaRNA, sin embargo la nomenclatura de microARN ha prevalecido. No queda claro por qué esta molécula fue catalogada como microARN en otras especies exceptuando humanos, si la pertenencia a un gen de scaRNA se encuentra conservada. En este contexto cabe preguntarse, ¿cómo se define a los microARNs? ¿Cuál es la división entre microARNs de biogénesis no canónica y otros ARNs pequeños del tipo ACA45 sRNA? Algunos autores utilizan como criterio la asociación a proteínas Ago, por ej. mediante experimentos de inmunoprecipitación<sup>139</sup>, aunque se han encontrado complejos de Ago2 con ARNt mitocondrial<sup>140</sup>, o el ejemplo de ACA45 sRNA. En un trabajo de revisión reciente, se ha realizado la distinción de aquellos microARNs de biogénesis no canónica que son transcriptos en tándem con o como parte de otro tipo de ARN pequeño, y se ha incluido en este grupo a ACA45 sRNA<sup>58</sup>. Sin embargo, esta molécula sigue sin figurar en miRBase, el catálogo de referencia para secuencias de microARNs.

*5) Los microARNs validados experimentalmente tienen una alta asociación con vías moleculares correspondientes a procesos metabólicos, especialmente procesos de biosíntesis de ácidos grasos.*

El patrón detectado por métodos bioinformáticos concuerda con el descrito a nivel de metaboloma y transcriptoma (de células únicas) de la fracción de LSC en pacientes con

LMC. En el trabajo en el que caracterizan el metaboloma, se demostró que las LSC dependen del metabolismo oxidativo en mayor medida que sus contrapartes más diferenciadas, y también con respecto a sus contrapartes normales. En este contexto, el tratamiento con un antibiótico que inhibe la síntesis de proteínas mitocondriales (tigeciclina) resultó efectivo para inhibir la capacidad *stem* y progenitora en muestras provenientes de pacientes con LMC, en ensayos *in vitro* e *in vivo*<sup>129</sup>. En paralelo, mediante secuenciación del transcriptoma de células únicas de pacientes con LMC al diagnóstico se detectó un enriquecimiento en genes pertenecientes a los *hallmarks* "metabolismo de ácidos grasos", "adipogénesis", "fosforilación oxidativa" y "glucólisis" en la fracción de células primitivas leucémicas BCR-ABL<sup>+</sup> vs. BCR-ABL<sup>-</sup> <sup>96</sup>. Resulta llamativo que estas diferencias no pudieron ser detectadas en un análisis *bulk* realizado *in silico* con los mismos datos, sugiriendo que el fenotipo alterado no es homogéneo, sino que corresponde probablemente a una subpoblación particular (o más de una). En leucemia mieloide aguda, se demostró que las LSC obtenidas de pacientes al diagnóstico son más dependientes del metabolismo de aminoácidos para la fosforilación oxidativa y supervivencia<sup>141</sup>. En algunos tumores sólidos como cáncer pancreático también se observó una mayor dependencia con la vía de fosforilación oxidativa en la fracción de células *stem* tumorales<sup>142</sup>. Estos resultados contrastan con la mayor dependencia de las células tumorales con la vía glicolítica, sugiriendo que se trata de un proceso específico de esta fracción primitivas<sup>143</sup>. Evidencias experimentales que vinculan el metabolismo de lípidos con la capacidad de autorrenovación de las HSC sugieren que ambos procesos podrían estar vinculados en las LSC<sup>144</sup>. En esta línea, en un modelo murino de LMC, se demostró que Alox5, enzima que cataliza la transformación de ácidos grasos en leucotrienos, es una reguladora crítica de la capacidad funcional de las LSC, pero no de sus contrapartes normales<sup>145</sup>.

## DISCUSIÓN FINAL Y PERSPECTIVAS FUTURAS

El primer objetivo de este trabajo de tesis fue evaluar si la cuantificación de los niveles de carga leucémica en la fracción de células primitivas y progenitoras podría tener un impacto clínico para la identificación de pacientes con distinta respuesta a la terapia con ITK. Elegimos estudiar a la fracción de células primitivas ya que el paradigma actual en LMC es que las LSC son las iniciadoras de la enfermedad, responsables del mantenimiento a largo plazo del clon leucémico, y posibles causantes de la recaída luego de la discontinuación del tratamiento. Los resultados obtenidos demostraron que los niveles de carga leucémica en dicha población, y también en la fracción de células progenitoras, disminuyen *in vivo* con el tratamiento con ITK, oponiéndose a la visión clásica de que las LSC son insensibles al efecto anti-proliferativo/pro-apoptótico de los mismos. Sin embargo, hay que tener en consideración que los ensayos LTC-IC no son equivalentes a los ensayos *in vivo*, considerados los ensayos modelo para evaluar la funcionalidad de las células madre hematopoyéticas.

A los 6 meses de tratamiento, se detectó una menor carga leucémica en la fracción primitiva en los pacientes que presentaron respuesta óptima al tratamiento con ITK en comparación con aquellos que presentaron signos de advertencia o falla. En este subgrupo, los signos de advertencia aparecieron rápidamente, en la mayoría de ellos, antes del año de tratamiento. Ésto implica que el ensayo LTC-IC, el cual requiere de mucho tiempo para la obtención de los resultados (aproximadamente 2 meses), no sería compatible con la rutina clínica. Sin embargo, los resultados sugieren que las fracciones primitivas y progenitoras reflejan (o generan) lo que ocurre en células más diferenciadas, lo cual implicaría que los mecanismos que operan en ambos casos son los mismos. Es decir, un paciente que presenta una buena respuesta en sangre periférica y MO al tratamiento con ITK, se esperaría que también presente una buena respuesta en la fracción de células primitivas y progenitoras.

La evaluación mediante esta metodología, de aquellos pacientes en RM profunda, evidenció la baja sensibilidad del método, siendo que la frecuencia y/o los niveles de expresión de *BCR-ABL1* en la fracción primitiva y progenitora fueron menores en estos pacientes o resultaron indetectables. Dado que uno de nuestros objetivos principales era poder cuantificar los niveles de carga leucémica en la fracción primitiva de pacientes candidatos a discontinuación del tratamiento, los cuales presentan respuestas moleculares profundas, sostenidas en el tiempo, la metodología utilizada no resultó satisfactoria. Es por ello que pusimos a punto la detección del clon leucémico a nivel de ADN genómico, ya que así nos independizaríamos de los niveles del ARNm de *BCR-ABL1*. Para ello implementamos una estrategia previamente publicada para la amplificación de la región correspondiente al punto de fusión en el cromosoma Philadelphia, seguida de secuenciación masiva del amplicón. Logramos optimizar la metodología y así diseñar primers específicos para el seguimiento de los pacientes individuales, los cuales se utilizaron para la detección a nivel de ADN genómico de colonias BCR-ABL1+ en ensayos LTC-IC. Este método, sin embargo, tiene algunas limitaciones:

-la región del punto de fusión *BCR-ABL1* es una región genómica enriquecida en regiones repetidas, secuencias palindrómicas, regiones de micro-homología, y bajo

contenido de GC en *ABL1*, que dificultan el diseño de primers óptimos y la obtención de una PCR de alta eficiencia y sensibilidad;

-la posibilidad de monitorear a los pacientes candidatos a discontinuar el tratamiento depende de que se haya guardado una muestra de ADN al diagnóstico, de buena calidad, para poder realizar la secuenciación del punto de fusión de forma retrospectiva;

-la frecuencia de células *BCR-ABL1*<sup>+</sup> es baja en pacientes con RM profunda, por lo que es necesario realizar el screening de una gran cantidad de colonias.

Este año se ha iniciado el primer ensayo clínico de discontinuación de ITK en Argentina (AST, Argentina STOP trial), y como grupo de trabajo participante, tendremos acceso a muestras de sangre periférica de estos pacientes, antes y después de la discontinuación del tratamiento. En este contexto, la posibilidad de evaluar la frecuencia de células primitivas *BCR-ABL1*<sup>+</sup> a nivel genómico es de gran interés para poner a prueba la hipótesis de que sus niveles están asociados a la posibilidad de discontinuar de forma segura. Al día de hoy, no se cuenta con marcadores pronósticos que permitan dividir a los pacientes en grupos con mayor o menor probabilidad de alcanzar el TFR. En este contexto, la cantidad de LSC residuales previo a la discontinuación, podría ser un factor a tener en cuenta. Las evidencias que hay hasta el momento apoyan, sin embargo, a la hipótesis alternativa, aquella que sostiene que la presencia de LSC es independiente de la posibilidad de alcanzar el TFR. Entre estas evidencias se encuentran:

-la detección de LSC a nivel de transcriptos *BCR-ABL1* mediante ensayos LTC-IC, en pacientes que lograron mantener el TFR, y la no detección de LSC en pacientes que presentaron recaída molecular luego de la discontinuación<sup>146</sup>. Estos resultados son muy relevantes, sin embargo, se pueden plantear dos objeciones: el ensayo LTC-IC realizado fue de 5 semanas en vez de 6 semanas, con lo que los autores evaluaron una fracción menos primitiva<sup>83</sup>; la detección a nivel de transcriptos tiene las limitaciones antes descriptas, con lo cual en los pacientes en los que no se detectaron LSC, tal vez se podrían haber detectado a nivel de ADN genómico;

-la detección del clon leucémico a nivel de ADN genómico en pacientes que discontinuaron y lograron el TFR<sup>147</sup>; lo más llamativo de los resultados obtenidos en este trabajo, es que los niveles de carga leucémica continuaron disminuyendo una vez suspendido el tratamiento, como así también en aquellos que presentaron recaída molecular y retomaron la medicación. La consideración que aplica en este caso es que la cuantificación se realizó en sangre entera, y no en la fracción de LSC.

Estos resultados sugieren que existen otros factores, independientes de las LSC, que influyen en la posibilidad de discontinuar el tratamiento y alcanzar el TFR. En particular, la participación del sistema inmunológico es una de las posibilidades que más interés ha despertado, dado que existen algunas evidencias que apoyan esta hipótesis. Entre ellas, se detectó que aquellos pacientes que lograron el TFR tenían, al momento de la discontinuación, más células Natural Killer (NK) maduras en SP que expresaban CD57 y CD16, en comparación con los que presentaron recaída molecular en los primeros 6 meses de discontinuar el IM<sup>148</sup>. Otros autores reportaron que las células dendríticas plasmacitoides que expresaban CD86 (ligando del checkpoint inmunológico CTLA-4)

estaban asociadas con la presencia de células T CD8<sup>+</sup> con fenotipo exhausto, y con una mayor probabilidad de recaída molecular<sup>149</sup>. En este contexto, sería de interés evaluar si la combinación de parámetros inmunológicos con los niveles de LSC a nivel de ADN genómico podría funcionar como factor pronóstico de alcanzar el TFR.

Otra hipótesis es que, dada la baja frecuencia con la cual las LSC se replican, es necesario que transcurra mucho tiempo para la erradicación del clon leucémico<sup>150,151</sup>. Dadas las evidencias de que las LSC tienen menor capacidad "stem" que sus contrapartes normales, es posible que el clon leucémico se agote antes que el reservorio de HSC, lo cual aseguraría el restablecimiento a largo plazo de la hematopoyesis normal. En este sentido, los resultados obtenidos en esta tesis apoyan dicha hipótesis. De ser así, sin embargo, no se sabe qué es lo que determina que, en algunos pacientes, el clon leucémico se agote antes que en otros. Sería de interés evaluar si existen factores extrínsecos, presentes en el nicho de la MO, capaces de regular la capacidad de autorrenovación de las LSC, y que estos factores difieran en aquellos pacientes que logran el TFR vs. aquellos que no.

Dadas las diferencias observadas en la capacidad de autorrenovación de las LSC, nos propusimos explorar los mecanismos moleculares que podían estar mediando dichas diferencias, a través del estudio del patrón de expresión de microARNs. Elegimos estudiar el perfil de expresión de microARNs ya que, por un lado, han sido poco explorados en la fracción de LSC de pacientes con LMC; y por el otro, se trata de un sistema de menor dimensionalidad que el estudio de los ARNm o proteínas. En este sentido, trabajos de otros autores han demostrado que los perfiles de microARNs pueden ser igual o más informativos que los de ARNm. Por ejemplo, en el trabajo de Lu *et al*, la evaluación global de los niveles de microARNs permitió la distinción entre tumores de distinto origen, y logró mejores agrupamientos de las muestras que el perfil de ARNm<sup>132</sup>.

El desarrollo de un organismo multicelular requiere la implementación de muchos mecanismos de control para poder definir (y mantener) patrones precisos de expresión génica en distintos tejidos, tipos celulares, y estadios de desarrollo. La robustez se puede definir como la capacidad de un sistema de mantener su función a pesar de recibir perturbaciones internas o externas. En este sentido, algunos autores han propuesto que la regulación mediada por microARNs otorga robustez a un proceso biológico, mediante la reducción de la variabilidad célula-célula ("ruido") en la expresión de ciertos genes en un contexto determinado, por ej. en una región del embrión durante el desarrollo<sup>152</sup>. Es decir, los microARNs, que suelen ser menos abundantes que los ARNm, tal vez no sean responsables de "encender" o "apagar" un programa transcripcional determinado, pero a través de la regulación fina, post-transcripcional, de los niveles de ARNm y/o proteínas, pueden contribuir a delimitar y mantener estables dichos programas de expresión génica específicos. En este contexto, la observación de que los genes con expresión específica de tejido tienden a tener ARNm cuyas regiones 3'UTR son más largas, con más sitios de unión a microARNs, que aquellos transcritos de expresión ubicua<sup>153</sup>, sugiere la participación de los microARNs en la definición de patrones específicos de expresión.

Los resultados obtenidos a partir de esta caracterización evidenciaron:

- una disminución global de los niveles de microARNs en la fracción de LSC y HSC de pacientes en comparación con las HSC de donantes sanos;

-tanto en la comparación de LSC vs. HSC de pacientes, como en la de HSC de pacientes vs. HSC de dadores sanos, se observó el agrupamiento de los microARNs subexpresados y sobreexpresados según las vías moleculares asociadas a sus posibles blancos, sugiriendo mecanismos conjuntos de desregulación;

-un enriquecimiento en vías asociadas a distintos procesos metabólicos en las LSC y HSC de pacientes, a través de microARNs que fueron validados por RT-qPCR;

-un enriquecimiento en vías asociadas a interacción receptor-matriz extracelular, proteoglicanos en cáncer y vía Hippo en las LSC y HSC de pacientes;

-un enriquecimiento en vías proliferativas en las HSC de pacientes, sugiriendo un fenotipo alterado a pesar de no albergar la mutación *BCR-ABL1*;

-la disminución en los niveles de un conjunto de microARNs y snoARNs localizados en la región 14q.32, susceptible de sufrir imprinting, en las LSC con respecto a HSC de dadores sanos.

A partir de estos resultados surgieron nuevas preguntas:

1) Dadas las evidencias recientes de otros autores de que la fracción de LSC de pacientes con LMC presenta un metabolismo oxidativo aumentado<sup>129</sup>, ¿podrían los microARNs desregulados ser mediadores de este mecanismo? De ser así, ¿la desregulación está relacionada a la actividad quinasa de BCR-ABL1? Como perspectivas futuras, evaluaremos la relevancia funcional de dichos microARNs en modelos *in vitro*. Los primeros ensayos funcionales que realizaremos serán ensayos de susceptibilidad al tratamiento con tigeciclina (antibiótico que inhibe la síntesis de proteínas mitocondriales), y la medición de parámetros asociados al metabolismo oxidativo (como tasa de consumo de oxígeno mitocondrial), en células en las que hayan sido inhibidos los microARNs de interés.

2) Respecto a la observación de que las HSC de pacientes presentan un perfil de expresión de microARNs desregulado, ¿cuál es el rol del microambiente en dicha regulación? ¿qué ocurre en esta fracción durante el tratamiento con ITK? Al ser progresivamente eliminadas las células leucémicas, ¿se restauran los niveles de microARNs en las HSC?

3) Respecto al patrón de disminución de los niveles de microARNs pertenecientes al *cluster* 14q.32 detectado en la fracción de LSC, ¿es común a otros pacientes con LMC? ¿Existe algún mecanismo epigenético asociado a dicha desregulación? ¿cuáles son las consecuencias funcionales de esta disminución?

Por último, los resultados obtenidos en esta tesis, aunque fueron realizados con muestras de pacientes con LMC, podría ocurrir que sean informativos de los procesos que ocurren en la fracción de células primitivas en otras patologías. En las LSC de leucemia mieloide aguda, por ejemplo, también se ha reportado un patrón metabólico alterado, particularmente dependiente de los aminoácidos para la fosforilación oxidativa<sup>141</sup>. Sería de interés explorar si los microARNs validados en nuestro sistema también están desregulados en dicha patología.

## REFERENCIAS

1. Rowley, J. A new consistent chromosomal abnormality in Chronic Myelogenous Leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature* **243**, 290–293 (1973).
2. Hughes, T. P., Ross, D. M. & Melo, J. V. *Handbook of Chronic Myeloid Leukemia*. (Springer International Publishing Switzerland, 2016). doi:10.1007/978-3-319-08350-6\_1
3. Guidelines, C. P. Chronic myeloid leukaemia : ESMO Clinical Practice Guidelines for diagnosis , treatment and follow-up † Clinical Practice Guidelines. **28**, 41–51 (2017).
4. Hughes, T. P., Ross, D. M. & Melo, J. V. *Handbook of Chronic Myeloid Leukemia*. (SpringerNature, 2016).
5. Mendizabal, A. M., Garcia-Gonzalez, P. & Levine, P. H. Regional variations in age at diagnosis and overall survival among patients with chronic myeloid leukemia from low and middle income countries. *Cancer Epidemiol.* (2013). doi:10.1016/j.canep.2013.01.002
6. Rohrbacher, M. & Hasford, J. Epidemiology of chronic myeloid leukaemia ( CML ). *Best Pract. Res. Clin. Haematol.* **22**, 295–302 (2009).
7. Bizzozero, O. J. *et al.* Radiation-Related Leukemia in Hiroshima and Nagasaki, 1946–1964. *N. Engl. J. Med.* **274**, 1095–1101 (1966).
8. Ross, D. *et al.* Distribution of genomic breakpoints in chronic myeloid leukemia : analysis of 308 patients. *Leukemia* **27**, 2105–2107 (2013).
9. Gabert, J. *et al.* Standardization and quality control studies of ‘real-time’ quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia – A Europe Against Cancer Program . *Leukemia* **17**, 2318–2357 (2003).
10. Beligoy, L. *et al.* Leucemia mieloide crónica. *Soc. Argentina Hematol. Guías diagnóstico y Trat.* 429–446 (2017).
11. Sokal, B. J. E. *et al.* Prognostic discrimination in ‘good risk’ chronic granulocytic leukemia. *Blood* **63**, 789–799 (1984).
12. Melo, J. V & Deininger, M. W. N. Biology of chronic myelogenous leukemia — signaling pathways of initiation and transformation. *Hematol. Oncol. Clin. North Am.* **18**, 545–568 (2004).
13. Druker, B. *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat. Med.* **2**, 561–566 (1996).
14. Kantarjian, H. M. *et al.* Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N. Engl. J. Med.* **346**, 645–652 (2002).
15. Hochhaus, A. *et al.* Long-term outcomes of Imatinib treatment for Chronic myeloid leukemia. *N. Engl. J. Med.* **376**, 917–927 (2017).
16. Hughes, T. P. *et al.* Frequency of Major Molecular Responses to Imatinib or Interferon Alfa plus Cytarabine in Newly Diagnosed Chronic Myeloid Leukemia. *N. Engl. J. Med.* **349**, 1423–1432 (2003).
17. Holyoake, T. L. & Vetrie, D. The chronic myeloid leukemia stem cell: Stemming the tide of persistence. *Blood* **129**, 1595–1606 (2017).
18. Shah, N. P. Overriding Imatinib Resistance with a Novel ABL Kinase Inhibitor. *Science (80- )*. **305**, 399–401 (2004).
19. Weisberg, E. *et al.* Characterization of AMN107 , a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **7**, 129–141 (2005).
20. Weisberg, E., Manley, P. W., Cowan-Jacob, S. W., Hochhaus, A. & Griffin, J. D. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nat. Rev. Cancer* **7**, 345–56 (2007).
21. Soverini, S., Benedittis, C. De, Mancini, M. & Martinelli, G. Mutations in the BCR-ABL1 Kinase Domain and Elsewhere in Chronic Myeloid Leukemia. *Clin. Lymphoma, Myeloma Leuk.* **15**, S120–S128 (2015).
22. O’Hare, T., Zabriskie, M. S., Eiring, A. M. & Deininger, M. W. Pushing the limits of targeted therapy in chronic myeloid leukaemia. *Nat Rev Cancer* **12**, 513–526 (2012).
23. Branford, S. *et al.* Rationale for the recommendations for harmonizing current methodology for detecting BCR-ABL transcripts in patients with chronic myeloid leukaemia. 1925–1930 (2006). doi:10.1038/sj.leu.2404388
24. Baccarani, M. *et al.* European LeukemiaNet recommendations for the management of chronic myeloid leukemia : 2013. *Blood* **122**, 872–884 (2013).
25. Hughes, T. P. *et al.* Impact of early dose intensity on cytogenetic and molecular responses in chronic- phase CML patients receiving 600 mg / day of imatinib as initial therapy. *Blood* **112**, 3965–3974 (2008).
26. Cross, N. C. P. *et al.* Development and evaluation of a secondary reference panel for BCR-ABL1

- quantification on the International Scale. *Leukemia* **30**, 1844–52 (2016).
27. Ruiz, M. S. *et al.* Standardization of molecular monitoring for chronic myeloid leukemia in Latin America using locally produced secondary cellular calibrators. *Leukemia* **30**, 2258–2260 (2016).
  28. Branford, S. *et al.* BCR-ABL Messenger RNA Levels Continue to Decline in Patients with Chronic Phase Chronic Myeloid Leukemia Treated with Imatinib for More Than 5 Years and Approximately Half of All First-Line Treated Patients Have Stable Undetectable BCR-ABL Using Strict Sens. *Clin. cancer Res.* **13**, 7080–7086 (2007).
  29. Mahon, F.-X. *et al.* Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. *Lancet Oncol.* **11**, 1029–35 (2010).
  30. Saußebe, S., Richter, J., Hochhaus, A. & Mahon, F. X. The concept of treatment-free remission in chronic myeloid leukemia. *Leukemia* **30**, 1638–1647 (2016).
  31. Fleming, H. E. *et al.* Wnt Signaling in the Niche Enforces Hematopoietic Stem Cell Quiescence and Is Necessary to Preserve Self-Renewal In Vivo. *Cell Stem Cell* **2**, 274–283 (2008).
  32. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
  33. Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441**, 518–522 (2006).
  34. Mckenzie, J. L., Gan, O. I., Doedens, M., Wang, J. C. Y. & Dick, J. E. Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. *Nat. Immunol.* **7**, 1225–1233 (2006).
  35. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis : A Human Perspective. *Cell Stem Cell* **10**, 120–136 (2012).
  36. Foudi, A. *et al.* Defining hematopoietic stem and progenitor cell turnover by analysis of histone 2B-GFP dilution. *Nat. Biotechnol.* **27**, 84–90 (2010).
  37. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells in vivo. *Blood* **117**, 4460–4466 (2011).
  38. Levin, R. H. *et al.* Persistent Mitosis of Transfused Homologous Leukocytes in Children Receiving Antileukemic Therapy. *Science (80-. )*. **142**, 1305–1311 (1963).
  39. Huntly, B. J. P. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587–596 (2004).
  40. Koschmieder, S. *et al.* Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis. *Blood* **105**, 324–334 (2005).
  41. Takahashi, B. N., Miura, I., Saitoh, K. & Miura, A. B. Lineage Involvement of Stem Cells Bearing the Philadelphia Chromosome in Chronic Myeloid Leukemia in the Chronic Phase as Shown by a Combination of Fluorescence-Activated Cell Sorting and Fluorescence In Situ Hybridization. *Blood* **92**, 4758–4763 (1998).
  42. Foley, S. B. *et al.* Expression of BCR / ABL p210 from a Knockin Allele Enhances Bone Marrow Engraftment without Inducing Neoplasia. *Cell Rep.* **5**, 51–60 (2013).
  43. Ichimaru, M., Tomonaga, M., Amenomori, T. & Matsuo, T. Atomic bomb and leukemia. *J. Radiat. Res.* **32 Suppl 2**, 14–19 (1991).
  44. Bose, B. S., Deininger, M., Gora-tybor, J., Goldman, J. M. & Melo, J. V. The Presence of Typical and Atypical BCR-ABL Fusion Genes in Leukocytes of Normal Individuals: Biologic Significance and Implications for the Assessment of Minimal Residual Disease. *Blood* **92**, 3362–3367 (1998).
  45. Nguyen, L. V, Vanner, R., Dirks, P. & Eaves, C. J. Cancer stem cells: an evolving concept. *Nat. Rev. cancer* **12**, 133–43 (2012).
  46. Valent, P. *et al.* Cancer stem cell definitions and terminology : the devil is in the details. *Nat. Rev. Cancer* **12**, 767–75 (2012).
  47. Chomel, J. *et al.* BCR-ABL expression in leukemic progenitors and primitive stem cells of patients with chronic myeloid leukemia. *Blood* **119**, 2964–2965 (2012).
  48. Chu, S. *et al.* Persistence of leukemia stem cells in chronic myelogenous leukemia patients in prolonged remission with imatinib treatment. *Blood* **118**, 5565–5572 (2011).
  49. Branford, S. *et al.* BCR-ABL1 doubling times more reliably assess the dynamics of CML relapse compared with the BCR-ABL1 fold rise : implications for monitoring and management. *Blood* **119**, 4264–4272 (2012).
  50. Hamilton, A. *et al.* Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival. *Blood* **119**, 1501–10 (2012).
  51. Corbin, A. S. *et al.* Human chronic myeloid leukemia stem cells are insensitive to imatinib

- despite inhibition of BCR-ABL activity. *J. Clin. Invest.* **121**, 396–409 (2011).
52. Jiang, X. *et al.* Chronic myeloid leukemia stem cells possess multiple unique features of resistance to BCR-ABL targeted therapies. *Leukemia* **21**, 926–935 (2007).
  53. Graham, S. M. *et al.* Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. *Blood* **99**, 319–325 (2002).
  54. Zhang, B. *et al.* Microenvironmental protection of CML stem and progenitor cells from tyrosine kinase inhibitors through N-cadherin and Wnt- $\beta$ -catenin signaling. *Blood* **121**, 1824–1838 (2013).
  55. Kumari, A., Brendel, C., Hochhaus, A., Neubauer, A. & Burchert, A. Low BCR-ABL expression levels in hematopoietic precursor cells enable persistence of chronic myeloid leukemia under imatinib. *Blood* **119**, 530–539 (2012).
  56. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
  57. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
  58. Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).
  59. Bartel, D. P. Review MicroRNAs : Target Recognition and Regulatory Functions. 215–233 (2009). doi:10.1016/j.cell.2009.01.002
  60. Liu, B., Shyr, Y., Cai, J. & Liu, Q. Interplay between miRNAs and host genes and their role in cancer. *Brief. Funct. Genomics* (2019). doi:10.1093/bfpg/elz002
  61. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
  62. Kim, D. H., Sætrom, P., Snøve, O. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230–16235 (2008).
  63. Ørom, U. A., Nielsen, F. C. & Lund, A. H. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and Enhances Their Translation. *Mol. Cell* **30**, 460–471 (2008).
  64. Calin, G. a & Croce, C. M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6**, 857–66 (2006).
  65. Guo, S. *et al.* MicroRNA miR-125a controls hematopoietic stem cell number. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14229–34 (2010).
  66. Peng, Y. & Croce, C. M. The role of MicroRNAs in human cancer. *Signal Transduct. Target. Ther.* **1**, (2016).
  67. Chomel, J. *et al.* Leukemic stem cell persistence in chronic myeloid leukemia patients with sustained undetectable molecular residual disease. *Blood* **118**, 3657–3660 (2011).
  68. Beillard, E. *et al.* Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using ‘ real-time ’ quantitative reverse-transcriptase polymerase chain reaction ( RQ-PCR ) – a Europe against cancer program. *Leukemia* **17**, 2474–2486 (2003).
  69. Vitsios, D. M. & Enright, A. J. Sequence analysis Chimira : analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics* **31**, 3365–3367 (2015).
  70. Feng, J. *et al.* GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* **28**, 2782–2788 (2012).
  71. Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W. & Rajewsky, N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
  72. Vlachos, I. S. *et al.* DIANA-miRPath v3 . 0 : deciphering microRNA function with experimental support. *Nucl* **43**, 460–466 (2015).
  73. Su, E. C. *et al.* ChemiRs : a web application for microRNAs and chemicals. *BMC Bioinformatics* 1–10 (2016). doi:10.1186/s12859-016-1002-0
  74. Hinske, L. C. G., Galante, P. A. F., Kuo, W. P. & Ohno-machado, L. A potential role for intragenic miRNAs on their hosts ’ interactome. *BMC Genomics* **11**, (2010).
  75. Ross, D. M. *et al.* Patients with chronic myeloid leukemia who maintain a complete molecular response after stopping imatinib treatment have evidence of persistent leukemia by DNA PCR. *Leukemia* **24**, 1719–1724 (2010).
  76. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, (2013).
  77. Wang, J. *et al.* CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* **8**, 652–654 (2011).
  78. Chen, K. *et al.* BreakDancer: an algorithm for high resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681 (2009).
  79. Untergasser, A. *et al.* Primer3Plus , an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, 71–74 (2007).

80. Chen, C. *et al.* Real-time quantification of microRNAs by stem – loop RT – PCR. *Nucleic Acids Res.* **33**, e179 1-9 (2005).
81. Di Rienzo, J. A. *et al.* InfoStat. (2011).
82. Coulombel, L. Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. *Oncogene* **23**, 7210–7222 (2004).
83. Sloma, I. *et al.* Genotypic and functional diversity of phenotypically defined primitive hematopoietic cells in patients with chronic myeloid leukemia. *Exp. Hematol.* **41**, 837–847 (2013).
84. Linhartova, J. *et al.* Characterization of 46 patient-specific BCR-ABL1 fusions and detection of SNPs upstream and downstream the breakpoints in chronic myeloid leukemia using next generation sequencing. 1–5 (2015). doi:10.1186/s12943-015-0363-8
85. Abel, H. J. *et al.* Detection of Gene Rearrangements in Targeted Clinical Next-Generation Sequencing. *J. Mol. Diagnostics* **16**, 405–417 (2014).
86. Ross, D. M. Minimal Residual Disease in Chronic Myeloid Leukaemia After Imatinib Treatment. (University of Adelaide, 2009).
87. Udomsakdi, C. *et al.* Rapid decline of chronic myeloid leukemic cells in long-term culture due to a defect at the leukemic stem cell level. *Proc Natl Acad Sci U S A* **89**, 6192–6196 (1992).
88. Bruns, I. *et al.* The hematopoietic stem cell in chronic phase CML is characterized by a transcriptional profile resembling normal myeloid progenitor cells and reflecting loss of quiescence. *Leukemia* **23**, 892–899 (2009).
89. Saussele, S. *et al.* Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. *Lancet Oncol.* **19**, 747–757 (2018).
90. Pagani, I. S. *et al.* BCR-ABL1 genomic DNA PCR response kinetics during first-line imatinib treatment of chronic myeloid leukemia. *Haematologica* **103**, 2026–2032 (2018).
91. Burchert, A. Maintaining low BCR-ABL signaling output to restrict CML progression and enable persistence. *Curr. Hematol. Malig. Rep.* **9**, 9–16 (2014).
92. Järås, M. *et al.* Isolation and killing of candidate chronic myeloid leukemia stem cells by antibody targeting of IL-1 receptor accessory protein. **107**, 16280–16285 (2010).
93. Gerber, J. M. *et al.* Genome-wide comparison of the transcriptomes of highly enriched normal and chronic myeloid leukemia stem and progenitor cell populations. *Oncotarget* **4**, 715–728 (2013).
94. Herrmann, H. *et al.* Dipeptidylpeptidase IV (CD26) defines leukemic stem cells (LSC) in chronic myeloid leukemia. *Blood* **123**, 3951–62 (2014).
95. Abraham, S. A. *et al.* Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells. *Nature* **534**, 341–346 (2016).
96. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23**, 692–702 (2017).
97. Warfvinge, R. *et al.* Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML. *Blood* **129**, 2384–2395 (2017).
98. Janssen, J. J. W. M. *et al.* Residual normal stem cells can be detected in newly diagnosed chronic myeloid leukemia patients by a new flow cytometric approach and predict for optimal response to imatinib. *Leukemia* **26**, 977–84 (2012).
99. Culen, M. *et al.* Quantitative assessment of the CD26+ leukemic stem cell compartment in chronic myeloid leukemia: Patient-subgroups, prognostic impact, and technical aspects. *Oncotarget* **7**, 33016–33024 (2016).
100. Quail, M. A. *et al.* A tale of three next generation sequencing platforms : comparison of Ion Torrent , Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 1 (2012).
101. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
102. Lopez, J. P. *et al.* Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med. Genomics* **8**, 35 (2015).
103. Zhao, S., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS One* **9**, e78644 (2014).
104. Lu, C. *et al.* Elucidation of the Small RNA Component of the Transcriptome. *Science (80-. )*. **309**, 1567–1569 (2005).
105. Morin, R. D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).
106. Kim, Y.-K., Yeo, J., Kim, B., Ha, M. & Narry Kim, V. Short Structured RNAs with Low GC Content Are Selectively Lost during Extraction from a Small Number of Cells. *Mol. Cell* **46**,

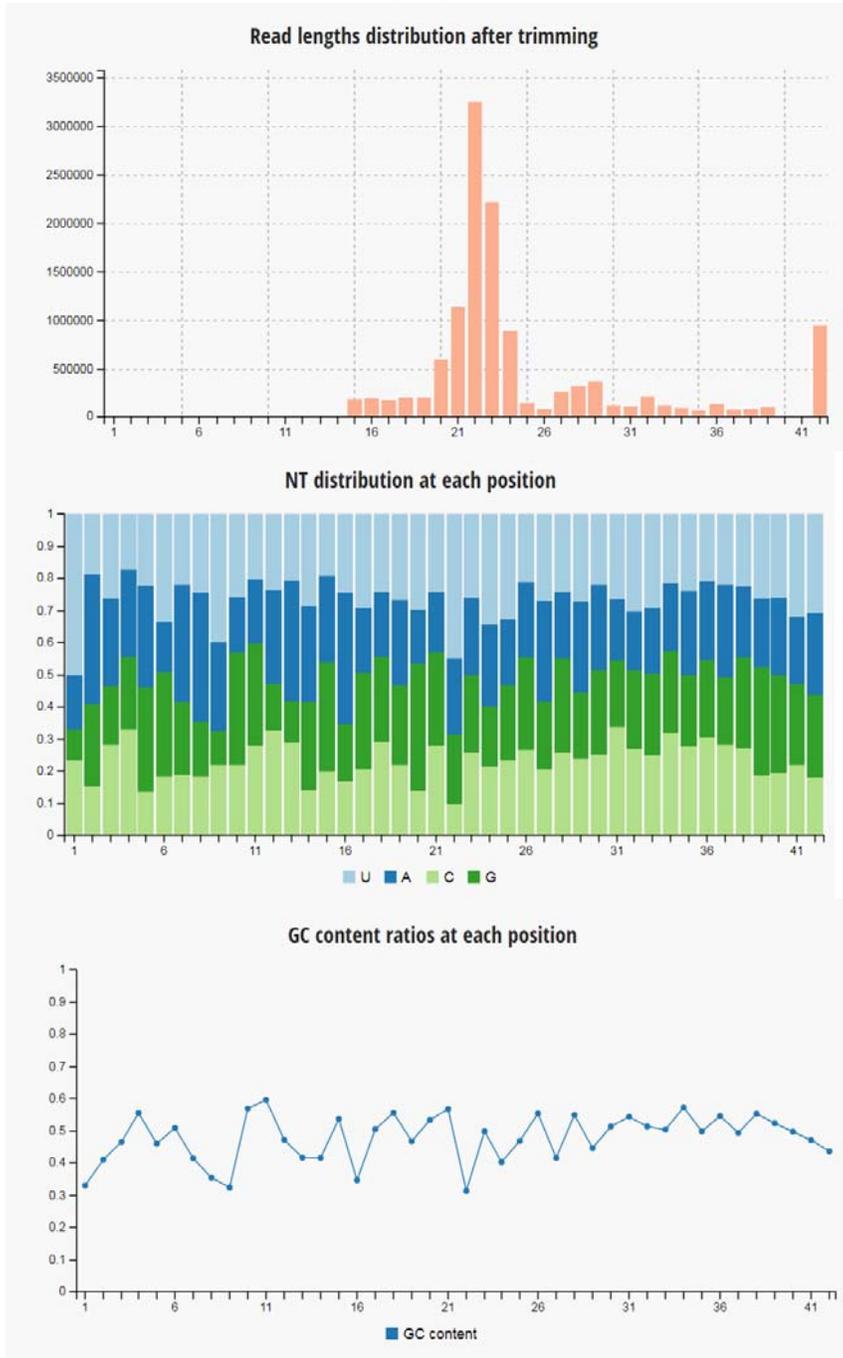
- 893–895 (2012).
107. Linsen, S. E. *et al.* Limitations and possibilities of small RNA digital gene expression profiling. *Nat. Methods* **6**, 474–476 (2009).
  108. Tian, G. *et al.* Sequencing bias : comparison of different protocols of MicroRNA library construction. *BMC Biotechnol.* **10**, (2010).
  109. Baran-gale, J. *et al.* Addressing Bias in Small RNA Library Preparation for Sequencing : A New Protocol Recovers MicroRNAs that Evade Capture by Current Methods. *Front. Genet.* **6**, 1–9 (2015).
  110. Agirre, X. *et al.* Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth. *Mol. Cancer Res.* **6**, 1830–40 (2008).
  111. Mehta, A. *et al.* The MicroRNA-132 and MicroRNA-212 Cluster Regulates Hematopoietic Stem Cell Maintenance and Survival with Age by Buffering FOXO3 Expression. *Immunity* **42**, 1021–1032 (2015).
  112. Tehler, D., Lund, A. H., Høyland-Kroghsbo, N. M. & Lund, A. H. The miR-10 microRNA precursor family. *RNA Biol.* **8**, 728–734 (2011).
  113. Potenza, N., Panella, M., Castiello, F., Mosca, N. & Amendola, E. Molecular mechanisms governing microRNA-125a expression in human hepatocellular carcinoma cells. *Sci. Rep.* 1–9 (2017). doi:10.1038/s41598-017-11418-3
  114. Naidu, S. *et al.* PDGFR-modulated miR-23b cluster and miR-125a-5p suppress lung tumorigenesis by targeting multiple components of KRAS and NF-κB pathways. *Sci. reportsq* **7**, 1–14 (2017).
  115. Wang, S. *et al.* The Endothelial-Specific MicroRNA miR-126 Governs Vascular Integrity and Angiogenesis. *Dev. Cell* **15**, 261–271 (2008).
  116. Li, X. *et al.* Transcriptional regulation of miR-10a / b by TWIST-1 in myelodysplastic syndromes. *Haematologica* **98**, 414–419 (2013).
  117. Shore, S. *et al.* Small RNA Library Preparation Method for Next-Generation Sequencing Using Chemical Modifications to Prevent Adapter Dimer Formation. *PLoS One* **11**, 1–26 (2016).
  118. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, 1–12 (2011).
  119. Peltier, H. J. & Latham, G. J. Normalization of microRNA expression levels in quantitative RT-PCR assays : Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* **14**, 844–852 (2008).
  120. Mestdagh, P. *et al.* A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* **10**, R64 (2009).
  121. Garzon, R. *et al.* Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proc. Natl. Acad. Sci.* **105**, 3945–3950 (2008).
  122. Sengupta, S. *et al.* MicroRNA 92b Controls the G1 / S Checkpoint Gene p57 in Human Embryonic Stem Cells. *Stem Cells* **27**, 1524–1528 (2009).
  123. Lechman, E. R. *et al.* Article Attenuation of miR-126 Activity Expands HSC In Vivo without Exhaustion. *Stem Cell* **11**, 799–811 (2012).
  124. Zhang, B. *et al.* Bone marrow niche trafficking of miR-126 controls the self-renewal of leukemia stem cells in chronic myelogenous leukemia. *Nat. Med.* **24**, 450–462 (2018).
  125. Benetatos, L. *et al.* The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cell. Mol. Life Sci.* **70**, 795–814 (2013).
  126. Nadal, E. *et al.* A MicroRNA Cluster at 14q32 Drives Aggressive Lung Adenocarcinoma. *Clin. cancer Res.* **20**, 3107–3117 (2014).
  127. Lechman, E. R. *et al.* MiR-126 Regulates Distinct Self-Renewal Outcomes in Normal and Malignant Hematopoietic Stem Cells. *Cancer Cell* **29**, 214–228 (2016).
  128. Chou, C. H. *et al.* MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302 (2018).
  129. Kuntz, E. M. *et al.* Targeting mitochondrial oxidative phosphorylation eradicates therapy-resistant chronic myeloid leukemia stem cells. *Nat. Med.* **23**, 1234–1240 (2017).
  130. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
  131. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* **153**, 654–665 (2013).
  132. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
  133. Mori, M. *et al.* Hippo Signaling Regulates Microprocessor and Links Cell-Density-Dependent miRNA Biogenesis to Cancer. *Cell* **156**, 893–906 (2014).

134. Qian, P. *et al.* The Dlk1-Gtl2 Locus Preserves LT-HSC Function by Inhibiting the PI3K-mTOR Pathway to Restrict Mitochondrial Metabolism. *Cell Stem Cell* **18**, 214–228 (2016).
135. Naka, K. *et al.* TGF- $\beta$  – FOXO signalling maintains leukaemia- initiating cells in chronic myeloid leukaemia. *Nature* **463**, 676–680 (2010).
136. Ronchetti, D. *et al.* The expression pattern of small nucleolar and small Cajal body-specific RNAs characterizes distinct molecular subtypes of multiple myeloma. *Blood Cancer J.* **2**, (2012).
137. Warner, W. A. *et al.* Expression profiling of snoRNAs in normal hematopoiesis and AML. *Blood Adv.* **2**, 151–163 (2018).
138. Ender, C. *et al.* A Human snoRNA with MicroRNA-Like Functions. *Mol. Cell* **32**, 519–528 (2008).
139. Goff, L. A. *et al.* Ago2 Immunoprecipitation Identifies Predicted MicroRNAs in Human Embryonic Stem Cells and Neural Precursors. *PLoS One* **4**, e7192 (2009).
140. Maniatakis, E. & Mourelatos, Z. Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. *RNA* **11**, 849–852 (2005).
141. Jones, C. L. *et al.* Inhibition of Amino Acid Metabolism Selectively Targets Human Leukemia Stem Cells. *Cancer Cell* **34**, 724–740 (2018).
142. Sancho, P. *et al.* MYC/PGC-1 $\alpha$  Balance Determines the Metabolic Phenotype and Plasticity of Pancreatic Cancer Stem Cells. *Cell Metab.* **22**, 590–605 (2015).
143. Pavlova, N. N. & Thompson, C. B. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.* **23**, 27–47 (2016).
144. Ito, K. *et al.* A PML-PPAR- $\delta$  pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nat. Med.* **18**, 1350–1358 (2012).
145. Chen, Y., Hu, Y., Zhang, H., Peng, C. & Li, S. Loss of the Alox5 gene impairs leukemia stem cells and prevents chronic myeloid leukemia. *Nat. Genet.* **41**, 783–792 (2009).
146. Chomel, J. C. *et al.* Leukemic stem cell persistence in chronic myeloid leukemia patients in deep molecular response induced by tyrosine kinase inhibitors and the impact of therapy discontinuation. *Oncotarget* **7**, 35293–35301 (2016).
147. Ross, D. M. *et al.* Long-term treatment-free remission of chronic myeloid leukemia with falling levels of residual leukemic cells. *Leukemia* **32**, 2572–2579 (2018).
148. Ilander, M. *et al.* Increased proportion of mature NK cells is associated with successful imatinib discontinuation in chronic myeloid leukemia. *Leukemia* **31**, 1108–1116 (2017).
149. Schütz, C. *et al.* Expression of the CTLA-4 ligand CD86 on plasmacytoid dendritic cells (pDC) predicts risk of disease recurrence after treatment discontinuation in CML. *Leukemia* **31**, 829 (2017).
150. Stein, A. M. *et al.* BCR – ABL Transcript Dynamics Support the Hypothesis That Leukemic Stem Cells Are Reduced during Imatinib Treatment. *Clin. cancer Res.* **17**, 6812–6822 (2011).
151. Roeder, I. *et al.* Dynamic modeling of imatinib-treated chronic myeloid leukemia : functional insights and clinical implications. *Nat. Med.* **12**, 1181–1184 (2006).
152. Ebert, M. S. & Sharp, P. A. Review Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* **149**, 515–524 (2012).
153. Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3'UTR Evolution. *Cell* **123**, 1133–1146 (2005).

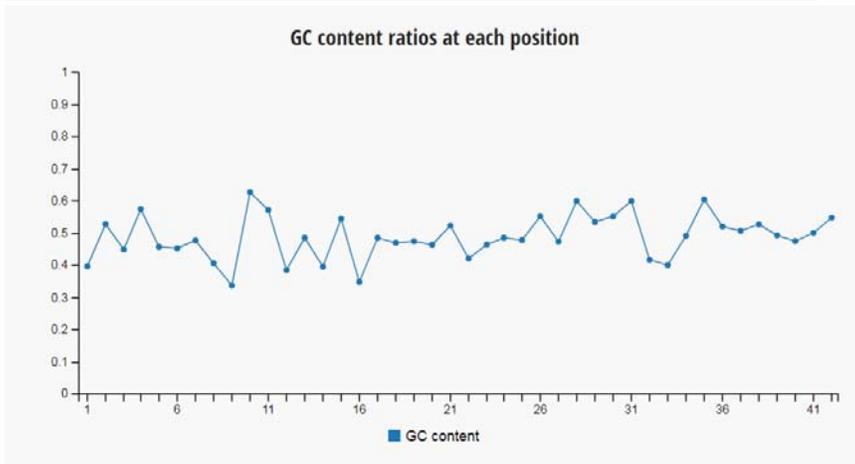
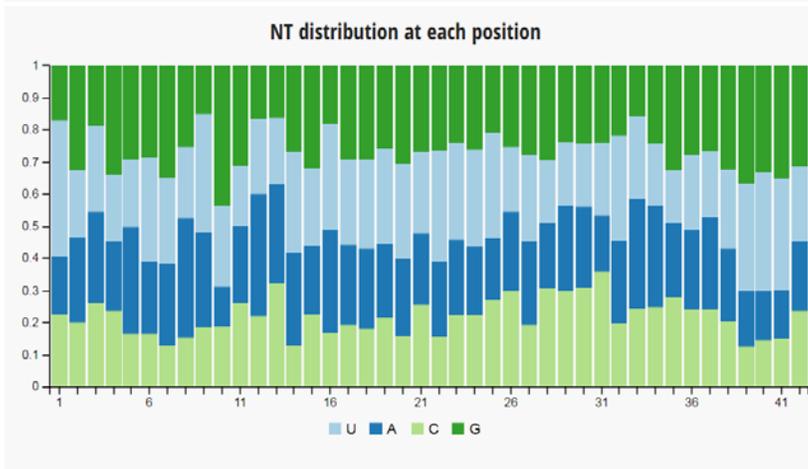
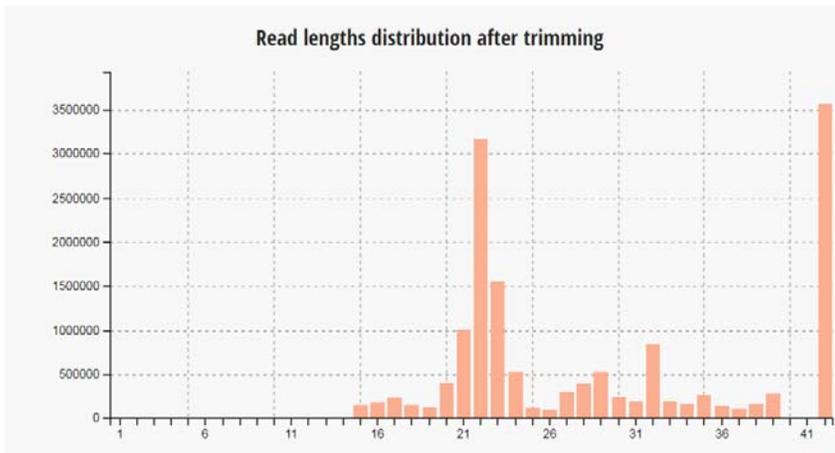
# Anexo I. Secuenciación masiva de la fracción de ARNs pequeños (TruSeq-Illumina HiSeq 2500)

## 1) Gráficas de calidad de las secuencias

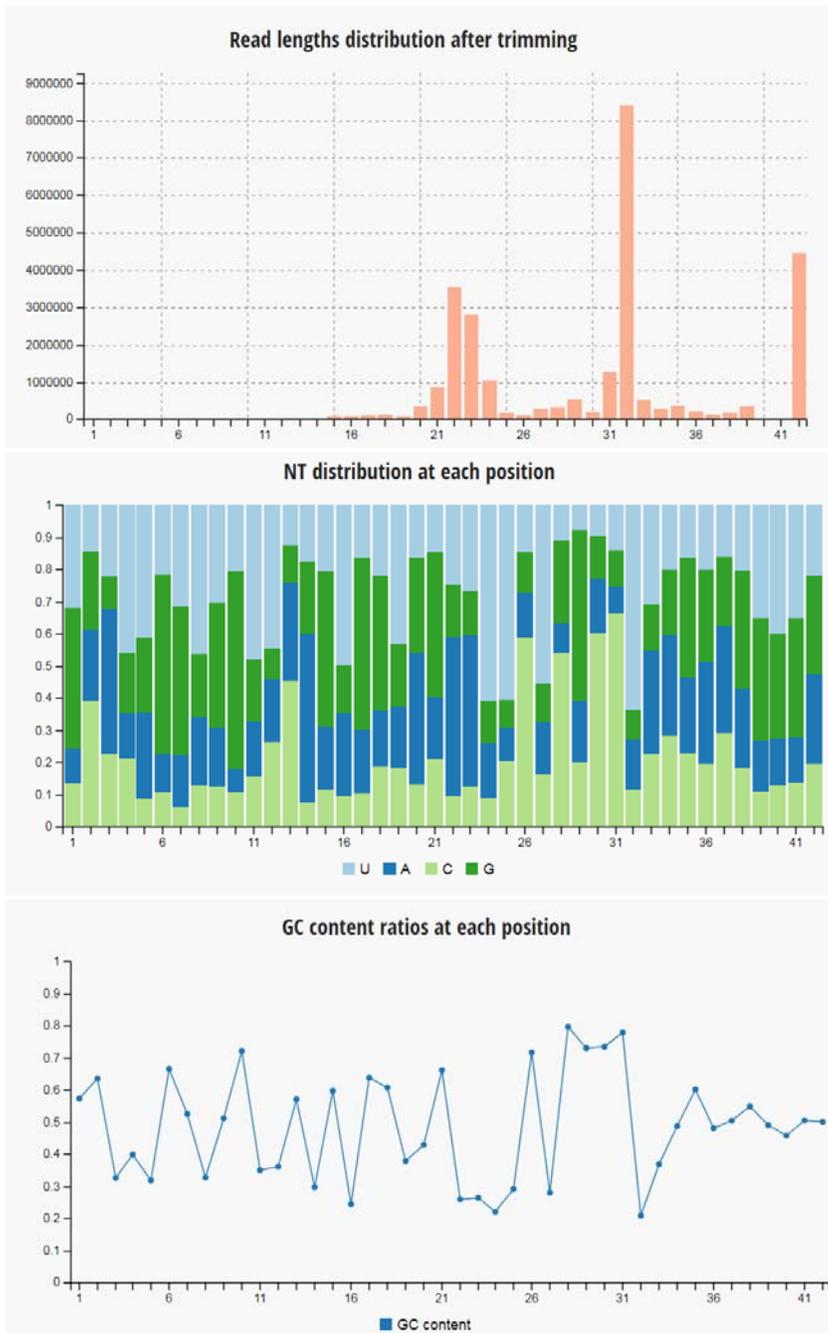
Pool 2 LMC CD34+ CD38- CD26+ (LSC):



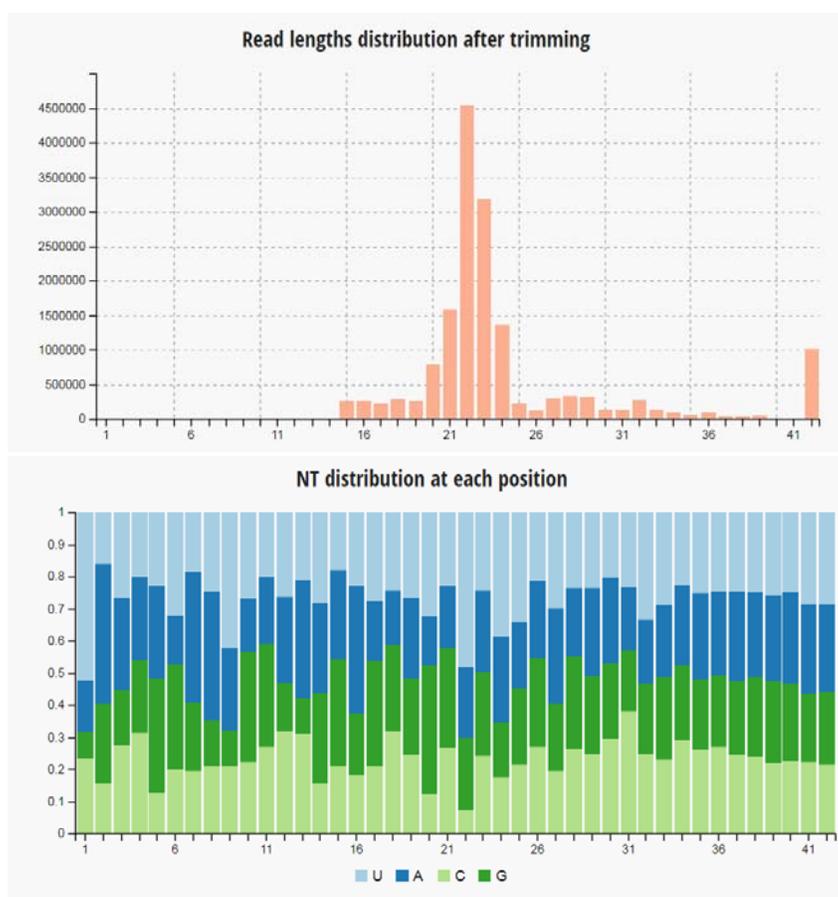
Pool 2 LMC CD34+ CD38- CD26- (HSC):



Pool dadoses sanos, CD34+ CD38+ (Progenitores):



Pool dadores sanos, CD34+ CD38-/dim (HSC):



**2) Profundidad de secuenciación**

	<b>Tamaño</b>	<b>No. lecturas totales (microARNs conocidos)</b>
<b>Pool 2, LMC, LSC</b>	200,8 MB	7.805.354
<b>Pool 2, LMC, HSC</b>	270,29 MB	6.276.876
<b>Pool DS, Progenitores</b>	441,53 MB	8.454.981
<b>Pool DS, HSC</b>	262,11 MB	11.233.176

**3) Análisis de microARNs novel por miRDeep2**

	<b>LSC (LMC)</b>	<b>HSC (LMC)</b>	<b>HSC (DS)</b>	<b>Progenitores (DS)</b>
<b># precursores totales</b>	587	230	711	532
<b># Novel</b>	223	86	282	187
<b>score promedio (precursores conocidos)</b>	7117	5519	8751	7458
<b>score promedio (novel)</b>	3,9	15	7,2	4,5
<b># novel (score mayor a 1)</b>	32	12	43	25

	Score miRDeep2				Conteos				Localización genómica	
	LMC		DS		LMC		DS		Crom.	Heb.
	LSC	HSC	Prog	HSC	LSC	HSC	Prog	HSC		
<b>Novel_miR_1</b>	1,26	1,26	1,26	1,26	141	59	191	150	chr5	+
<b>Novel_miR_2</b>	1,45	1,45	1,45	1,45	40	18	17	22	chr2	+
<b>Novel_miR_3</b>	1029	1512	904	1465	1459	2623	1483	2128	chr15	+

## Anexo II. Secuencias de *primers* y sondas

<i>Primers</i> y sondas para la detección de <i>BCR-ABL1</i> y <i>ABL1</i> por RT-qPCR	
ABL1 Fw (ENF1003)	GGAGATAAACTCTAAGCATAAACTAAAGGT
ABL1 Rv (ENR1063)	GATGTAGTTGCTTGGGACCCA
BCR-ABL1 p210 Fw (ENF501)	TCCGCTGACCATCAAYAAGGA
BCR-ABL1 p210 Rv (ENR561)	CACTCAGACCCTGAGGCTCAA
Sonda ABL1 (ENPr1043)	CCATTTTTGGTTTGGGCTTCACACCATT
Sonda BCR-ABL1 p210 (ENP541)	CCCTTCAGCGCCAGTAGCATCTGA

<i>Primers stem-loop</i> para la retrotranscripción de microARNs	
hsa-miR-92b-3p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGAGGC
hsa-miR-125a-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCACAG
hsa-miR-182-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGTGT G
novel-3 RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAAGAC
hsa-miR-2355-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTGTC
hsa-miR-126-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCGCGTA
hsa-miR-132-3p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCGACCA
hsa-let-7a-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTTCTAT
hsa-miR-10a-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCACAAA
hsa-miR-196a-5p RT	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCCCAAC
<i>Primers</i> Fw para la qPCR	
hsa-miR-92b-3p Fw	TGCATTTCTATTGCACTCGTC
hsa-miR-125a-5p Fw	TGATCCCTGAGACCCTTTAAC
hsa-miR-182-5p Fw	GGTTTGGCAATGGTAGAACT
novel-3 Fw	CGGACTGAAGGTAGATAGAACAG

hsa-miR-2355-5p Fw	TCGAAGATCCCCAGATACAAT
hsa-miR-126-5p Fw	CAGCGGCATTATTACTTTTGG
hsa-miR-132-3p Fw	CTCGGTAACAGTCTACAGCCA
hsa-let-7a-5p Fw	GCGGTGAGGTAGTAGGTTGT
hsa-miR-10a-5p Fw	GGCTACCCTGTAGATCCGAA
hsa-miR-196a-5p Fw	GCGTCGTAGGTAGTTTCATGTT
snARN U6 Fw	GCTTCGGCAGCACATATACTAAAAT
<i>Primer Rv para la qPCR</i>	
Primer Rv universal miRNAs	GTGCAGGGTCCGAGGT
snARN U6 Rv	CGCTTCACGAATTTGCGTGT CAT

<i>Primers para PCR long range del punto de fusión BCR-ABL1 en ADN genómico</i>	
<i>Primers BCR Fw</i>	
BCRB2F (exón 13)	ACTCGTGTGTGAACTCCAGACT
BCRB3F (exón 14)	TGGGTTTCTGAATGTCATCGTCC
<i>Primers ABL1 Rv, sets multiplex</i>	
<b>Set A</b>	
ABL1AR	AACACGGAGAAGTGGCAAACCTC
ABL2AR	TAAGGCAGTTACCAGGAAGCATTT
ABL3AR	GGAAGGAGGAGGAAATGACAGCA
ABL4AR	ACTGTTCACTAAGTGGCACTGTG
ABL5AR	CATGATGTGCTTTGCAGGGTAGC
<b>Set B</b>	
ABL1BR	GATCTGAAGCACAAGCACGGTTC
ABL2BR	GGAACAGGAATCCTAATGGCCAAC
ABL3BR	ACATGGGGCACAGTCTCTTGATG
ABL4BR	AGGAATGGGTATGCTGGGGTTG
ABL5BR	ATGGAATGACTCCCACCTGAAAG
<b>Set C</b>	
ABL1CR	AAGCCACTGGCACACTTCATACG
ABL2CR	TGGACCAGGCTTTAGCCCTATCA
ABL3CR	CCCGCAGTATCCCTCAAAATCAG
ABL4CR	ACATGAGGTTTGCAGAAGCACCA
<b>Set D</b>	
ABL1DR	AGCCATAACCATTCTCCCAAGCA
ABL2DR	ACCAAAGCCTCCCTTGTACCTC
ABL3DR	TCATGGAGAAAGGGGAGAAGT
ABL4DR	CCAGAAGAACACCCCAAGAAGGA
ABL5DR	AAACATTCTGCCGCATCTGGATT

<i>Primers paciente-específicos para la detección de BCR-ABL1 en ADN genómico</i>		
		Tamaño estimado del producto de PCR
N22 Fw 2	GAACCTTATTACACTTCGAG	456pb
N22 Rv 2	AGTAGAGACAAAGTTTCACC	
N22 Fw	AGAGTTAGCTTGTCACCT	166pb
N22 Rv	CTCATTAGCTATAACTCTGC	
N38 Fw BCR_1	GCAGGTGGATCGAGTAATTG	235 pb
N38 Rv ABL	TGCTTGAGCCAAACAAATGG	
N38 Fw BCR_2	AGCAGATTTGGCTGCTCTGTC	113 pb
N38 Rv ABL	TGCTTGAGCCAAACAAATGG	
N35 Fw BCR	TCATGGACCCCAAACCTTGTTTC	103 pb
N35 Rv ABL	GGCCAGGTAAATTCAGACCAC	
N35 Fw BCR	TCATGGACCCCAAACCTTGTTTC	209 pb
N35 Rv ABL1	GCATTGACTTCTGCCAGCTAAC	
MP5920 Fw BCR_ext	TTAGGAGCAGTTTCTCCCTGAG	389 pb
MP5920 Rv ABL_ext	CGGTGAAGCCCCGTCTC	
MP5920 Fw BCR_int	AGCAGACGCTCCTCAGATG	253 pb
MP5920 Rv ABL_int	GCACCTGTAGTCCCAGCTACTC	
B-RAF-S Fw	TGCTTGCTCTGATAGGAAAATG	227 pb
B-RAF-S Rv	GCATCTCAGGGCCAAAAT	