

ESTUDIO COMPARATIVO DE SECUENCIAS DE ARN EN MUESTRAS DE MÉDULA ÓSEA DE LEUCEMIA OBTENIDAS DE MODELOS ANIMALES, PACIENTES EN ETAPA TEMPRANA Y PACIENTES BAJO TRATAMIENTO.

RESUMEN

La leucemia es una enfermedad con una alta ocurrencia en el Ecuador y en el mundo, ya que afecta a los diferentes tipos de células sanguíneas, impactando principalmente a los niños y adolescentes. En los últimos años, la investigación molecular de la leucemia ha sido basada en el estudio de la expresión genética, ya que con esta información se ha llegado a crear posibles tratamientos con terapia génica que proporcionen mejores efectos terapéuticos. Este proyecto se centró en realizar un análisis comparativo de las muestras de leucemia linfoblástica aguda, evaluando los efectos de tres tipos de muestras: 1) paciente con leucemia de células T sin tratar, 2) paciente con leucemia de células B tratado con prednisolona y 3) un modelo animal *knockout* con leucemia de células T. En la muestra tratada con prednisolona se observó una mayor cantidad de transcritos de genes relacionados con la función de P53, destacando la influencia del gen RPL11 en la ubiquitinación de MDM2 y la inhibición de la expresión de P53. En contraste, la muestra no tratada mostró una mayor cantidad de transcritos del gen WDR74 asociado a la regulación de MDM2. En el modelo animal *knockout* se identificó una mayor cantidad de transcritos del gen NOTCH 1 asociado está encargado de la regulación de señales celulares, aumentando la proliferación celular y la aparición de células sanguíneas inmaduras, lo cual promueve la leucemia. Estos hallazgos subrayan la complejidad de la leucemia y señalan la necesidad de enfoques terapéuticos más específicos. Sin embargo, se deben tener en cuenta limitaciones, como la falta de normalización de datos, los diferentes tipos de leucemias y especies, y el uso de una muestra para cada tipo, lo que impide obtener conclusiones definitivas sobre el proceso y destaca la necesidad de investigaciones adicionales.

Palabras clave: Leucemia, Prednisolona, Transcritos, RPL11, MDM, WDR74

ABSTRACT

Leukemia is a disease with a high occurrence in Ecuador and worldwide, as it affects different types of blood cells, mainly impacting children and adolescents. In recent years, molecular research on leukemia

has been based on the study of gene expression, as this information has led to the development of possible gene therapy treatments that provide better therapeutic effects. This project focused on conducting a comparative analysis of acute lymphoblastic leukemia samples, evaluating the effects of three types of samples: 1) untreated T-cell leukemia patient, 2) B-cell leukemia patient treated with prednisolone, and 3) a knockout animal model with T-cell leukemia. In the prednisolone-treated sample, a higher number of transcripts of genes related to the function of P53 was observed, highlighting the influence of the RPL11 gene on MDM2 ubiquitination and inhibition of P53 expression. In contrast, the untreated sample showed a higher number of transcripts of the WDR74 gene associated with MDM2 regulation. In the knockout animal model, a higher number of transcripts of the NOTCH1 gene associated with the regulation of cellular signals were identified, increasing cell proliferation and the appearance of immature blood cells, promoting leukemia. These findings underscore the complexity of leukemia and highlight the need for more specific therapeutic approaches. However, limitations such as lack of data normalization, different types of leukemias and species, and the use of one sample for each type must be considered, preventing definitive conclusions about the process and emphasizing the need for further research.

Keywords: Leukemia, Prednisolone, Transcripts, RPL11, MDM2, WDR74

INTRODUCCIÓN

La leucemia es una enfermedad que afecta a las células sanguíneas como los glóbulos blancos, rojos y plaquetas, las cuales tienen un crecimiento anormal y descontrolado, aumentando la susceptibilidad a contraer afecciones como la anemia, debilidad, sangrado, entre otros (Gutiérrez y Kentsis, 2018). La leucemia tiene una alta incidencia a nivel mundial, ya que en el 2023 se han reportado 59.610 nuevos casos de esta enfermedad, teniendo una mayor incidencia en personas entre 0 y 19 años (Siegel *et al.*, 2023). En Ecuador se ha reportado que la tasa de incidencia de la leucemia en niños es de 6,2 por cada 100.000 personas, siendo el principal caso de leucemia en niños y teniendo una tasa de mortalidad de 4,5 por cada 100.000 habitantes (Ministerio de Salud Pública del Ecuador, 2023).

La leucemia ha sido clasificada según varios parámetros como el tipo de células sanguíneas afectadas (linfoides o mieloides) y la velocidad de desarrollo de la enfermedad (aguda o crónica) (Sekeres *et al.*, 2020). Estos diferentes tipos de leucemia pueden afectar a personas de diversas edades y perfiles. Entre los tipos más comunes se encuentran la leucemia linfóide aguda (LLA), que afecta a las células linfoides

inmaduras en la médula ósea, y la leucemia mieloide aguda (LMA), caracterizada por el desarrollo anormal de células mieloides inmaduras. También están la leucemia linfocítica crónica (LLC), que se origina en las células linfocíticas maduras y suele desarrollarse lentamente, y la leucemia mieloide crónica (LMC), que afecta a las células mieloides maduras y progresa de manera más gradual (Chennamadhavuni *et al.*, 2023). Cada uno de estos tipos puede manifestarse en diferentes grupos de edad y perfiles de pacientes, lo que complica aún más el panorama de esta enfermedad.

Al ser la leucemia una enfermedad multifactorial, no solo por su componente genético, sino también por factores ambientales y de estilo de vida, el tipo de tratamiento puede variar. Esta variabilidad en los factores causales hace que el enfoque terapéutico deba ser adaptado cuidadosamente a las circunstancias de cada paciente. A pesar de que la quimioterapia y la radioterapia son tratamientos comunes para esta enfermedad, es importante tener en cuenta que la elección del tratamiento puede variar significativamente en función del tipo específico de leucemia y su etapa de desarrollo (Juliussen y Hough, 2016).

Por otra parte, es fundamental destacar que la leucemia no solo afecta la médula ósea, sino que también puede comprometer otros órganos vitales como el cerebro. Esto añade una capa adicional de complejidad al manejo clínico de la enfermedad y puede llevar a consecuencias graves como una muerte prematura. Por lo tanto, el trasplante de médula ósea se presenta como una de las opciones de tratamiento más efectivas en muchos casos de leucemia. Sin embargo, para que este procedimiento sea viable, es esencial que exista un alto grado de histocompatibilidad entre el donante y el receptor, generalmente con un porcentaje mayor al 80%, como han sugerido investigaciones recientes (Medina *et al.*, 2022). Esta alta compatibilidad es necesaria para minimizar el riesgo de rechazo del injerto y garantizar una recuperación exitosa del paciente.

En los últimos años, gracias a los avances en estudios moleculares, hemos mejorado nuestra comprensión del componente genético de la leucemia. Estos avances nos han permitido investigar con mayor detalle las anomalías cromosómicas, como las translocaciones e inversiones, que dan lugar a genes quiméricos anómalos como el BCR-ABL, conocido como el cromosoma Filadelfia en la LMC (Grimwade *et al.*, 2016). Además, las nuevas tecnologías, como la secuenciación de nueva generación, nos han permitido identificar las variantes patogénicas de genes que podrían estar relacionadas con la leucemia. Por ejemplo, se han descubierto mutaciones que afectan la interacción entre el gen PHF6 (*Plant Homeodomain-like Finger 6*) y PAF1 (*Polymerase-Associated Factor 1*), lo que altera la síntesis

ribosomal y conduce a una maduración incompleta de las células sanguíneas durante el proceso de hematopoyesis (Prada-Arismendy *et al.*, 2017).

Todos estos avances científicos han permitido entender con mayor profundidad la leucemia y el efecto que tienen los medicamentos en el tratamiento de esta enfermedad. Por ejemplo, la prednisolona, que es un glucocorticoide, tiene un efecto inhibitorio sobre la producción de citocinas, alterando la expresión de los oncogenes e induciendo la apoptosis celular (Inaba y Pui, 2010). Sin embargo, se ha evidenciado que puede ocasionar efectos secundarios, ya que inhibe el efecto del sistema inmunitario aumentando la probabilidad de infecciones, así como la función renal, ya que aumenta la presión arterial (Shah y Kumar, 2013). Sin embargo, para obtener una comprensión más profunda de la leucemia y sus tratamientos, se han utilizado tanto estudios *in vitro* como *in vivo*, incluyendo modelos de ratones *knockout*, los cuales son organismos genéticamente modificados, en donde se inhibe la expresión de ciertos genes. Estos modelos de ratones *knockout* han proporcionado valiosas perspectivas sobre cómo las mutaciones genéticas afectan el proceso de hematopoyesis y contribuyen a la formación de células sanguíneas anómalas (Mohanty y Heuser, 2021). Un ejemplo particular de interés ha sido el estudio del gen NRAS (*Neuroblastoma Rat Sarcoma*) y su influencia en la proliferación celular, así como en la inhibición de receptores de tirosina quinasa en células leucémicas, revelando que cuando NRAS está mutado existe una proliferación celular descontrolada (Liang *et al.*, 2006).

El presente proyecto tiene como objetivo realizar un estudio comparativo de transcritos utilizando datos de tres tipos de muestras: un paciente sin tratar con leucemia de células T, un paciente con LLA en células B tratado con prednisolona y un modelo animal (ratón *knockout* con leucemia de células T). El objetivo fue analizar los datos de RNA-seq de los tres tipos de muestras en busca de las diferencias de la abundancia de los transcritos entre las tres muestras estudiadas. En los resultados obtenidos dentro del proyecto se observó que en la muestra tratada existe la presencia del gen RPL11, involucrado en el control de la expresión de P53, además de la presencia de genes ribosomales como: RPS27, involucrado en el proceso de muerte celular, mientras en la muestra no tratada se observó una gran cantidad de transcritos pertenecientes al gen WDR74, que igual que RPL11, se encuentra regulando la actividad de P53 y finalmente en el modelo animal se obtuvo la presencia de NOTCH1 y PTMA. Los resultados obtenidos en el proyecto proporcionan conclusiones definitivas, ya que no se dio una normalización de los datos, lo cual conlleva a que las gráficas de la abundancia de los transcritos no tienen una misma escala, dificultando el análisis de las muestras.

METODOLOGÍA

Obtención de los datos

El set de datos utilizado para la realización de este proyecto fue obtenido a partir de los datos públicos en NCBI a partir de la base de datos GEO (Edgar *et al.*, 2002), que es una base en donde solo se colocan los datos de secuencias de ARN para la búsqueda de la información de las secuencias. Se utilizaron las palabras claves “Leukemia differential expression on bone marrow”, en donde se encontraron tres números de acceso correspondientes a estudios moleculares en LLA: SRR7281351 en el que los pacientes fueron tratados con prednisolona con LLA de células B (Autry *et al.*, 2020); SRR3196703 es un experimento en donde se estudiaron las características genómicas y transcriptómicas de muestras de células T en pacientes pediátricos con LLA (Hospital CHU Sainte-Justine, 2016); finalmente, SRR7526005, que corresponde a un estudio en ratones *knockout*, en donde se silenció el gen PHF6 para determinar los efectos que tiene en el desarrollo de células T inmaduras que ocasionan LLA (Wendorf *et al.*, 2019).

Para descargar las secuencias se utilizó la herramienta de SRA-Toolkit (NIH Genomics, 2022), siguiendo el tutorial proporcionado por el NCBI (<https://github.com/ncbi/sra-tools/wiki/02.-Installing-SRA-Toolkit>). En este estudio se instaló esta herramienta en una máquina virtual con Ubuntu 20.04 y para la instalación se empleó el código:

```
# Descarga de SRA toolkit:
wget http://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.4.1/sratoolkit.2.4.1-ubuntu64.tar.gz

# Extracción del archivo
tar xzvf sratoolkit.2.4.1-ubuntu64.tar.gz

# Uso de SRA-Toolkit
fastq-dump --dump --split-files SRR7281351
```

Con el código anterior se descargaron las secuencias *paired-end*, de plataforma Illumina, secuenciados con un *HiSeq 2500 System*.

Preparación de los datos

Para cada secuencia estudiada se procedió a realizar el análisis de calidad mediante FastQC v0.11.9 (Andrews, 2010), en donde se determinó la calidad de las secuencias, para posteriormente limpiarlas

con la herramienta Trimmomatic v.040 (Bolger *et al.*, 2014), la cual elimina los adaptadores y secuencias de baja calidad. Para este proceso se utilizó se empleó el siguiente comando:

```
# TRAILING:20: elimina las bases finales de baja calidad (con un Phred score menor a 20).
# MINLEN:50: descarta las secuencias más cortas de 50 bases después del recorte de cabezas y la
eliminación de las secuencias de baja calidad.
# HEADCROP:10: recorta las primeras 10 bases de cada secuencia, lo que puede ser útil para eliminar
adaptadores o secuencias no deseadas que pueden aparecer al inicio de la lectura.

trimmomatic PE /home/jupyter/compartida/PROYECTO\
FINAL/Datos/leucemia/paciente_leucemia.fastq /home/jupyter/compartida/PROYECTO\
FINAL/Datos/leucemia/paciente_leucemia_limpio.fastq TRAILING:20 MINLEN:50 HEADCROP:10
```

Una vez realizada la limpieza de las secuencias, se ejecutó nuevamente el análisis de control de calidad y se apreció la presencia de una gran cantidad de secuencias con N, lo que indicaba que eran bases que no pudieron ser asignadas en el proceso de secuenciación. Cada archivo descargado fue fragmentado en secuencias más pequeñas con Python, lo cual facilitó la manipulación de los datos, tomando en cuenta la estructura de los archivos Fastq y separando las secuencias completas. Finalmente, cada archivo dividido tuvo un peso de 350 MB, lo cual posibilitó la manipulación y análisis de los datos en un ambiente local.

Homologación de muestras

Para las secuencias obtenidas, una de cada tratamiento que se está analizando, debe existir un proceso previo de homologación de secuencias, ya que como se realizó un análisis comparativo entre diferentes especies, es necesario realizar un estudio de homología, para este propósito se utilizó la herramienta Megablast (Morgulis *et al.*, 2008), en la plataforma bioinformática de libre acceso Galaxy (Blankenberg *et al.*, 2011). Megablast emplea un algoritmo que fragmenta la secuencia en unidades más pequeñas denominadas "palabras", generalmente compuestas por 28 nucleótidos. Luego, realiza una búsqueda en la base de datos de secuencias para identificar aquellas palabras que coinciden con las de la secuencia de consulta. Seguidamente, lleva a cabo un alineamiento local entre la secuencia de consulta y las secuencias de la base de datos que contienen coincidencias de palabras. Para concluir, se realiza el cálculo de un puntaje de similitud para cada alineamiento, generando una lista de las secuencias de la base de datos que poseen los puntajes de similitud más elevados.

Pseudoalineamiento

Para el pseudoalineamiento de las secuencias estudiadas se utilizó kallisto v0.48.0 (Bray *et al.*, 2016), en donde se realizó la indexación de la secuencia de referencia, necesaria para realizar la cuantificación de cada uno de los transcritos a partir de los *k-mers*, para la cual se utilizó la función `quant` de kallisto, la cual genera una tabla de abundancias que contiene el nombre del transcrito (`TARGET_ID`), la longitud de la secuencia total (`LENGTH`), la estimación de la longitud de cobertura de las secuencias (`EST_LENGTH`), la estimación de los conteos (`EST_COUNTS`) y la abundancia relativa de los transcritos por millón (TPM). Para la realización de este paso se aplicó el siguiente comando:

Indexar la secuencia de referencia

```
kallisto index -i Homo_sapiens.GRCh38.cdna.all.idx Homo_sapiens.GRCh38.cdna.all.fa.gz
```

Donde *i* Indica el archivo que va a ser indexado

Cuantificación del transcrito

```
kallisto quant -i Homo_sapiens.GRCh38.cdna.all.idx -o output_directory_no_tratado_muestra1 --  
paired -l 200 -s 20 segmento_paciente_Leucemia_tratado_R1_segmento_1.fastq  
segmento_paciente_Leucemia_tratado_R2_segmento_1.fastq
```

Donde:

#-o output_directory_no_tratado_muestra1: Define el directorio de salida donde Kallisto almacenará los resultados de la cuantificación. En este caso, los resultados se guardarán en un directorio llamado "output_directory_no_tratado_muestra1".

#--paired: Indica que las lecturas de entrada son pareadas (paired-end). Esto significa que estás proporcionando dos archivos FASTQ, uno para las lecturas R1 (segmento_paciente_Leucemia_tratado_R1_segmento_1.fastq) y otro para las lecturas R2 (segmento_paciente_Leucemia_tratado_R2_segmento_1.fastq).

#-l 200: Especifica la longitud media de los fragmentos de ARN. Este valor es importante para el cálculo de las probabilidades de alineación y se utiliza en el modelo de estimación de abundancia.

#-s 20: Indica la desviación estándar de la longitud de los fragmentos de ARN. Al igual que el parámetro `-l`, este valor es utilizado en el modelo de estimación de abundancia.

Preparación de tabla de abundancia

A partir de las tablas de abundancia se preparó una tabla general que fue importada a R para el análisis de expresión diferencial con las herramientas `awk` v5.2.0 (Aho *et al.*, 1992) y `split` (Stallman y Granlund,

1997), con las cuales se filtraron estas tablas de abundancia, obteniéndose las columnas de importancia que son el nombre del gen, el conteo de lecturas que tenían estas secuencias y la abundancia en TPM. Para este propósito se utilizó el siguiente comando:

```
awk -F"\t" 'BEGIN{OFS="\t"} NR>1{split($1, gene, "."); split(gene[1], name, "_"); print name[3], name[4], $1, "tratados", $5, $6}' output_directory_tratado_muestra10.tsv > tratado.tsv
```

Donde:

F"\t": es la opción de awk para especificar el delimitador de campo, en este caso, el carácter de tabulación.

'BEGIN{OFS="\t"}: esta es la sección BEGIN del script awk que se ejecuta antes de procesar las líneas del archivo. En este caso, se establece el delimitador de campo de salida (OFS) como un carácter de tabulación.

NR>1: es una condición que se evalúa para cada línea del archivo. NR es una variable interna de awk que indica el número de registro (línea) actual. En este caso, se evalúa si el número de registro es mayor que 1, lo que significa que se omitirá la primera línea del archivo.

{split(\$1, gene, "."); split(gene[1], name, "_"); print name[3], name[4], \$1, "tratados", \$5, \$6}: este es el bloque principal del script awk que se ejecuta para cada línea que cumple con la condición anterior.

split(\$1, gene, "."): se divide el primer campo (\$1) utilizando el carácter punto como delimitador y se guarda en el arreglo gene.

split(gene[1], name, "_"): se divide el primer elemento del arreglo gene (que contiene el nombre del gen) utilizando el carácter subrayado como delimitador y se guarda en el arreglo name.

print name[3], name[4], \$1, "tratados", \$5, \$6: se imprime el tercer elemento del arreglo name, el cuarto elemento del arreglo name, el primer campo (\$1), la cadena "tratados" y los campos \$5 y \$6.

Comparación de abundancia de transcritos en R

Para el análisis diferencial se utilizó R debido a que, dentro de este lenguaje de programación de código abierto, existen librerías como EdgeR (Robinson, McCarthy y Smyth, 2010) que pertenece al paquete Bioconductor, el cual permite manipular los datos de secuenciación. En este estudio, además, se utilizó el conteo de los transcritos mediante el método RPKM (*Reads Per Kilobase per Million mapped reads*), ya que este método toma en cuenta el tamaño de la muestra y la longitud de las lecturas, mejorando la calidad de los datos obtenidos.


```
(base) jupyter@ububio:~/compartida/PROYECTO FINAL/Datos/tratados$ head leucemia_tratado_limpio.fastq
+SRR7281384.1 1 length=168
GCAACTGCAGGGCTCTCTTGCTTAGAGTGGTGGCCAGCGCCCTGCTGCCCGGGGGCACTGCAGGGCCCTCTAGAGGGCCCTGCAGTGCCCCGGGGCAGCAGGGGGCGCTGCCACCCTCT
AAGCAAGAGAGCCCTGCAGTTGCCCTAGTCGCC
+SRR7281384.1 1 length=168
```

Figura 3: Archivo resultante de Trimmomatic. Secuencias donde se eliminaron los adaptadores y las secuencias no especificadas por la plataforma de secuenciación.

Homologación de muestras

En la Tabla 1 se presenta el análisis de homología entre la muestra de ratón y el paciente tratado, en donde se aprecia que una similitud entre 86 a 98% con *e-values* inferiores a 1. Por su parte en la Tabla 2, se tiene el análisis entre los transcritos del paciente sin tratar y el modelo animal, se tiene una similitud entre 90 a 98 %, ya que las dos muestras estudiadas provienen del mismo tipo de leucemia. Estas comparaciones entre las muestras estudiadas proporcionan un panorama claro de la homología entre el transcriptoma del ser humano y del ratón, con lo cual se puede afirmar con fiabilidad que las muestras estudiadas tienen una similitud, brindando una base sólida para el análisis bioinformático en cuanto a la diferencia de especies.

MUESTRA DE RATÓN	MUESTRA PACIENTE TRATADO	PORCENTAJE DE SIMILITUD (%)	LONGITUD DE ALINEACIÓN	NÚMERO DE DESAJUSTES (mismatches)	GAPS	E-VALUE
SRR7526005.44	SRR728383.1507688	96	100	4	0	1.53e-39
SRR7526005.44	SRR728383.1507683	95,098	102	5	0	5.49e-39
SRR7526005.44	SRR728383.1507695	95,918	98	4	0	1.97e-38
SRR7526005.44	SRR728383.1507689	95,876	97	4	0	7.10e-38
SRR7526005.44	SRR728383.1507681	93,396	106	6	1	2.55e-37
SRR7526005.228	SRR7281384.1331743	89.583	192	7	3	1.15e-60
SRR7526005.228	SRR7281384.1331743	86.583	192	7	3	6.15e-33

Tabla 1: Salida de Megablast para los datos de ratón vs paciente con tratamiento.

MUESTRA PACIENTE SIN TRATAR	MUESTRA DE RATÓN	PORCENTAJE DE SIMILITUD (%)	LONGITUD DE ALINEACIÓN	NÚMERO DE DESAJUSTES (mismatches)	GAPS	E-VALUE
SRR3196703.2603.1	SR- R7526005.393220	98.24	50	1	0	1.16e-17
SRR3196703.2603.1	SR- R7526005.392319	98.35	50	1	0	1.16e-17
SRR3196703.3570.1	SR- R7526005.344610	90.36	50	6	0	4.20e-12
SRR3196703.2657.1	SR- R7526005.333310	90.118	51	2	1	2.16e-67

Tabla 2: Salida de Megablast para los datos de paciente sin tratamiento vs ratón

Obtención de tablas de abundancia

El comando kallisto quant se ejecutó para cada una de las muestras analizadas. En la Tabla 3 se observa la abundancia de expresión génica para la muestra que no ha sido sometida a tratamiento resultante de la ejecución de este comando. Los datos en esta tabla reflejan una amplia gama de valores de abundancia. Algunos genes muestran una abundancia relativamente alta, con valores superiores a 1.000, mientras que otros presentan conteos bajos, menores a uno. Esta variabilidad en los niveles de expresión indica la presencia de genes con una expresión génica tanto alta como baja en la muestra no tratada.

TARGET_ID	LENGT H	EFF_LENGTH	EST_COUNTS	TPM
ENST00000390469.2	520	321	16	258.648
ENST00000535880.2	425	226	1	229.608
ENST00000390344.2	552	353	3	441.003
ENST00000390345.2	688	489	42.707	453.195
ENST00000390346.2	537	338	1	153.525
ENST00000426402.2	544	345	17.293	260.103
ENST00000390477.2	720	521	8	796.796
ENST00000611116.2	978	779	330	219.822
ENST00000632774.1	1266	1067	159.409	775.254
ENST00000631539.1	2619	2420	43.062	0.923364
ENST00000633698.1	1678	1479	99	347.345
ENST00000626472.2	1485	1286	108.098	436.187
ENST00000390548.6	2619	2420	43.062	0.923364
ENST00000641136.1	2751	2552	502.403	102.156

Tabla 3: Tabla de abundancia relativa de expresión de los genes. Salida de kallisto quant para los datos de pacientes sin tratamiento.

Análisis comparativo de transcritos

- **Tratado y no tratado**

En la Figura 4 se presenta un gráfico de caja para cada uno de los tratamientos (no tratado y tratado), que permite visualizar la distribución de los datos en escala RPKM de cada muestra. Es importante destacar que la comparación directa entre los dos tratamientos puede ser limitada debido a la falta de normalización de los datos, lo cual dificulta la interpretación precisa de las diferencias observadas. Sin embargo, a pesar de esta limitación, se puede observar que en cada una de las distribuciones existen datos atípicos, con una cantidad posiblemente mayor en la muestra tratada, lo que podría indicar una alta expresión de ciertos transcritos.

- **No tratado y modelo animal**

Por otro lado, en la Figura 6 se encuentra un diagrama de caja que resalta la distribución de los transcritos en las muestras no tratadas y las muestras de ratones *knockout*. Aquí se aprecia claramente que en la distribución de la muestra no tratada existen varios *outliers*, lo que también se vio en la figura 4, además de un solo *outlier* para la muestra animal.

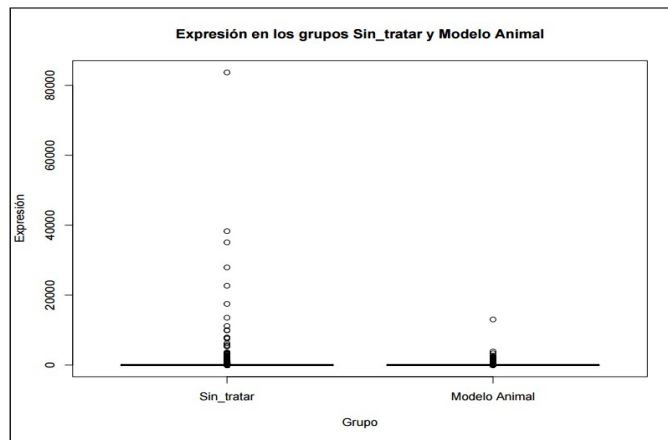


Figura 6: Diagrama de caja: sin tratar y modelo animal en escala de RPKM. Distribución del no tratado tiene transcritos *outlier*, mientras el modelo animal tiene datos más uniformes.

En la Figura 7, el gráfico de dispersión muestra la muestra del modelo animal en color rojo, mientras que las muestras no tratadas se presentan en azul. Dentro de las muestras no tratadas, es notable la presencia de un número significativo de transcritos de un gen específico, identificado como WDR74 (ENST00000538098.2). Además, en esta muestra se observa otro transcrito relevante como RPL13A (ENST00000474171.1). Por otro lado, en el modelo animal, se destaca el número de transcritos de dos genes de gran relevancia: PTMA (ENST00000409115.8) y NOTCH1 (ENST00000651671.1).

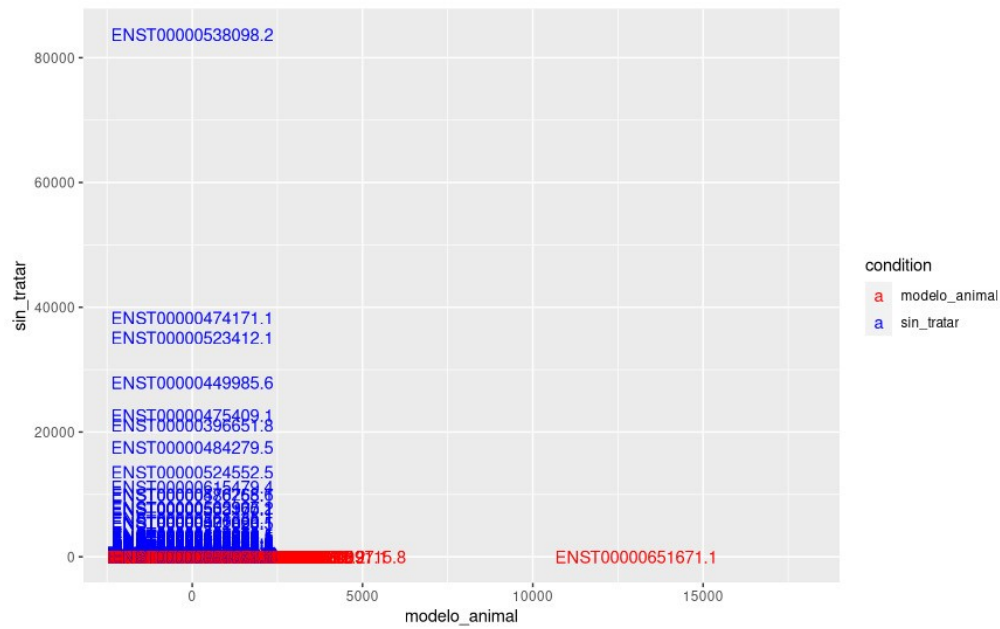


Figura 7: Gráfico de dispersión: muestras sin tratar y de modelo animal. En el modelo animal se ven transcritos involucrados con las subunidades ribosomales.

- **Muestra tratada con prednisolona y modelo animal**

En la Figura 8 se observa los diagramas de caja de las distribuciones de los datos del modelo animal y tratado con prednisolona. Se aprecia, que por la diferente escala usada en esta figura se tienen varios datos *outliers*, tanto en la muestra tratada como en el modelo animal.

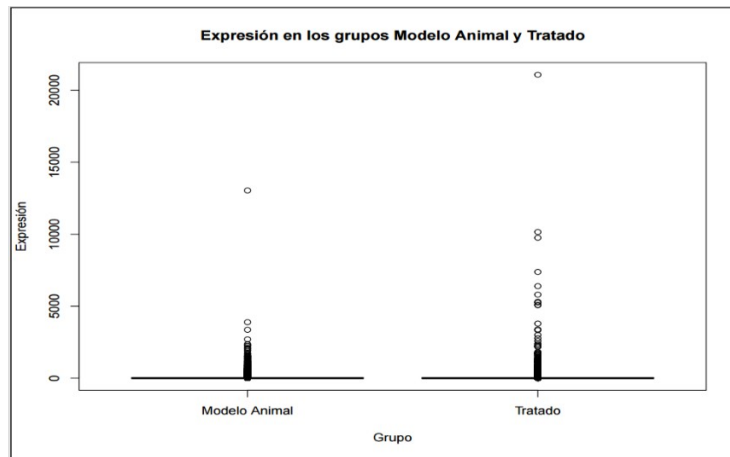


Figura 8: Diagrama de caja: modelo animal vs muestra tratada con prednisolona en escala de RPKM. Los dos tratamientos tienen transcritos outliers.

La Figura 9 indica la gráfica de dispersión entre el modelo animal de color rojo y la muestra tratada con prednisolona representada en color azul. En esta figura se destaca que en las dos muestras existen genes que poseen una gran cantidad de transcritos, tomando en cuenta que las escalas son diferentes. Por

ejemplo, en el modelo animal, sobresalen dos genes de gran importancia: NOTCH1 (ENST00000651671.1) y PTMA (ENST00000409115.8), mientras que en la muestra tratada se encuentran los genes RPL11 (ENST00000374550.8), RPS27 (ENST00000651669.1) y CD52 (ENST00000374213.3).

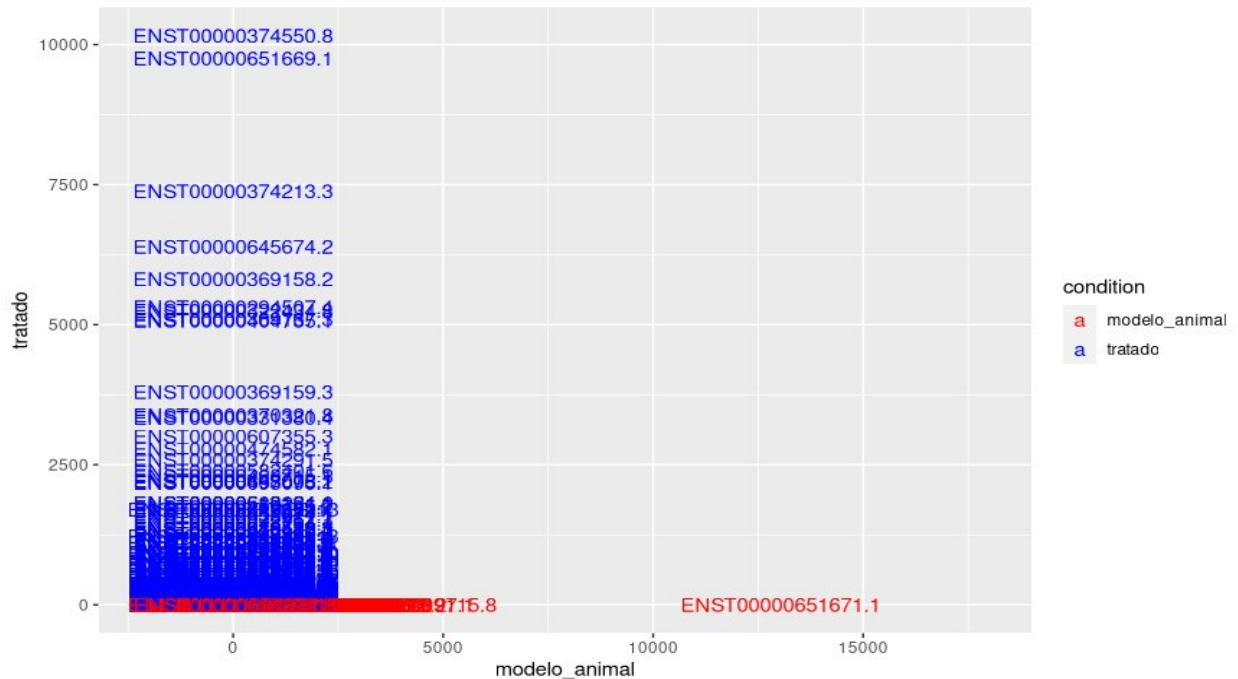


Figura 9: Gráfico de dispersión: muestras de modelo animal y tratada con prednisolona, muestras que tienen transcritos con valores altos.

DISCUSIÓN

La utilización de herramientas como la línea de comandos de bash de Linux es de vital importancia en el estudio de grandes volúmenes de datos como en el caso de la secuenciación de RNA-seq. Estos conjuntos de datos suelen tener un tamaño considerable, a menudo superior a los 5 GB, debido a la complejidad inherente de los organismos eucariotas que se están analizando. La secuenciación de RNA-seq genera una cantidad masiva de información, ya que se obtienen millones, e incluso miles de millones, de secuencias cortas que representan los fragmentos de ARN presentes en la muestra. El procesamiento y análisis de esta información requiere herramientas eficientes capaces de manejar grandes volúmenes de datos sin agotar la memoria RAM del equipo (Mora-Márquez *et al.*, 2021)

El análisis de homología que se realizó en esta investigación, arrojó una similitud de entre 86 a 98 % y entre la muestra de ratón y el paciente tratado, mientras que entre la del paciente sin tratar y el modelo animal, se tiene una similitud entre 90 a 98 %. Esta diferencia que existe entre las dos comparaciones, se debe a que, aunque son todas muestras de LLA, difieren en el tipo de célula afectada, Breschi, Gingeras y

Guido (2017) mencionan que los transcriptomas del ratón y del ser humano en general son conservadas, pero dependen del tipo de tejido y los genes específicos que se están estudiando. En consecuencia, aunque la similitud obtenida proporciona una base sólida para afirmar que las muestras estudiadas comparten características homólogas, es fundamental abordar con precaución los análisis posteriores. La variabilidad en el tamaño de los transcritos, influenciada por factores como la longitud de las secuencias, introduce complejidades adicionales en los estudios bioinformáticos. Este factor puede impactar la comparación de la cantidad de transcritos y requiere una consideración cuidadosa al interpretar los resultados y al realizar análisis más detallados

Kallisto quant es una herramienta de línea de comandos ampliamente utilizada en el procesamiento de datos de secuenciación de ARN (RNA-seq), ya que optimiza los recursos informáticos, utilizando el uso de k-mers, facilitando así un análisis rápido y preciso de la expresión génica en grandes conjuntos de datos transcriptómicos. Durante este proyecto se confirmó su fiabilidad, simplicidad y eficacia en el análisis de muestras de RNA-seq. Kallisto quant proporcionó tablas de abundancia basadas en TPM, lo que facilitó la evaluación de la cantidad de transcritos.

Es importante destacar que, en este proyecto no se implementó una normalización a escala logarítmica. Esta omisión presenta un desafío para llevar a cabo un análisis de expresión diferencial, ya que los datos obtenidos mediante RNA-seq siguen una distribución binomial. Es crucial subrayar que estos datos deben normalizarse a una escala común para facilitar un análisis preciso de la cantidad de expresión diferencial. La expresión de un gen puede variar significativamente según las condiciones biológicas de la muestra y el proceso de secuenciación. Puede haber una mayor abundancia de transcritos secuenciados de un gen en particular, lo que podría enmascarar la cantidad de transcritos de otros genes, generando así posibles falsos positivos (Evans, Hardin y Stoebel, 2018). Por lo tanto, esta investigación se limita únicamente a la comparación del número de transcritos presentes en cada muestra teniendo en cuenta las limitaciones asociadas.

Tampoco se puede ignorar que la muestra no tratada corresponde a un tipo de leucemia linfoblástica aguda de células B, mientras que la muestra de ratón y de no tratado son muestras de LLA de células T. Dadas las diferencias substanciales entre las LLA de células B y T, la interpretación de los efectos específicos de la prednisolona en la muestra tratada requiere una consideración cautelosa. La respuesta a la quimioterapia y otros tratamientos es una de las diferencias significativas entre la leucemia linfoblástica aguda (LLA) de células B y T (Li, et.al, 2022). Se ha observado que las células de LLA de células T tienden a ser más resistentes a los medicamentos quimioterapéuticos convencionales en

comparación con las células de LLA de células B (Barz, et.al, 2023). Esto puede influir en la eficacia del tratamiento y en el pronóstico de la enfermedad. Además, los pacientes con LLA de células B son más susceptibles a las terapias dirigidas disponibles, como los inhibidores de la tirosina quinasa de clase ABL para los casos de cromosoma Filadelfia positivo y algunos casos de cromosoma Filadelfia-like, así como a varias inmunoterapias. Por el contrario, las opciones de tratamiento novedosas para la LLA de células T son limitadas. (Teachey y Pui, 2019).

A pesar de las limitaciones analíticas de los datos, se observó que en la muestra tratada con prednisolona existen varios genes, como es el caso del gen RPL11 (Figuras 5 y 9), cuya expresión es notablemente mayor que los otros genes en la muestra. Esto se podría deber a que, como ha demostrado, cuando este gen tiene un alto número de ARN, hay mayores probabilidades de inhibir la acción del gen TP53, encargado de la senescencia celular, arresto del ciclo celular, apoptosis y reparación del ADN (Kim *et al.*, 2022). En condiciones normales, RPL11 promueve la ubiquitinación de la proteína MDM2 (Murine Double Minute 2), facilitando así la activación del gen TP53, que a su vez regula la proliferación celular (Dai *et al.*, 2006). En condiciones anormales, se ha descrito que RPL11 genera una gran cantidad de unidades ribosomales que se acumulan en el nucleosoma, generando estrés. Estas moléculas se unen a una mayor cantidad de MDM2, impidiendo la expresión de la proteína p53 y promoviendo la proliferación celular (Liu *et al.*, 2016). El número de transcritos de este gen tan altas debido a que esta muestra fue tratada con prednisolona y una posible actividad de este fármaco es que actúa a nivel celular activando los iones de Ca^{+2} del retículo endoplasmático al citosol, incrementando así la actividad del citocromo c y estimulando la apoptosis celular. Este efecto reduce la sobreexpresión del gen RPL11 causando la muerte celular (Inaba y Pui, 2010).

Además, en la muestra tratada se encontró la presencia del gen CD52 (ENST00000374213.3) (Figuras 5 y 9), el cual juega un papel muy importante en la regulación positiva de la concentración de iones calcio citosólico (National Center of Biotechnology, 2024), que influencia procesos de liberación celular y es crucial para muchas funciones celulares, incluyendo la señalización celular, la contracción muscular, la secreción de neurotransmisores y la activación de enzimas (Jiang *et al.*, 2009). El gen CD52 produce un antígeno, que se encuentra naturalmente en la superficie de linfocitos maduros y otras células del sistema inmunitario excepto en células plasmáticas, granulocitos, eritrocitos, plaquetas o células progenitoras hematopoyéticas, siendo un marcador molecular importante en el tratamiento de la Leucemia (Zhao *et al.*, 2017).

En la muestra tratada (Figuras 5 y 9) se puede apreciar la presencia del gen RPS27 (ENST00000651669.1), el cual juega un papel clave al codificar para la proteína ribosomal S27, componente esencial de la subunidad ribosomal pequeña 40S (Xiong *et al.*, 2018). Esta proteína no solo es vital para la síntesis de proteínas celulares y el procesamiento de los ARNr 18S, sino que también está implicada en una variedad de procesos biológicos fundamentales, como la regulación de la apoptosis, la progresión del ciclo celular y la proliferación celular. Además, se ha observado que las proteínas ribosómicas, cuando no están asociadas a ribosomas debido a alteraciones en la biogénesis ribosomal, responden a estímulos tanto extracelulares como intracelulares, desempeñando múltiples funciones extrarribosómicas. Estas incluyen la regulación de la estabilidad genómica, la transformación celular y la señalización inmunitaria, entre otros aspectos críticos para el desarrollo y la homeostasis celular (Xu *et al.*, 2016). En el contexto de la leucemia, se ha evidenciado que mutaciones sobre este gen, al igual que una sobreexpresión, puede generar daños en el funcionamiento de la médula ósea (Vlachos, 2017)

Por su parte, en la muestra no tratada (Figuras 5 y 7) se aprecia que el gen WDR74 (ENST00000538098.2) es el que tiene mayor cantidad de transcritos en esta muestra. Este gen es importante, ya que está involucrado en una variedad de procesos celulares críticos, como son la regulación de la división celular, la transcripción génica, la señalización transmembranal, la modificación del ARN mensajero y la fusión de vesículas (Liu *et al.*, 2018). Además, juega un papel insustituible en el ensamblaje del ribosoma 60S, que es crucial para la biogénesis de los ribosomas y la síntesis proteica en todas las células vivas. Este proceso es esencial para mantener la proliferación y el crecimiento de las células cancerosas y su alteración puede resultar en estrés ribosomal e inhibición de la proliferación celular (Li *et al.*, 2020). Aunque se han identificado asociaciones entre la desregulación de la biogénesis de los ribosomas y la progresión del cáncer, la comprensión del papel de WDR74 en el desarrollo del cáncer sigue siendo limitada. Se ha observado que WDR74 también regula el potencial metastásico de las células cancerosas a través de la actividad de ubiquitinación de MDM2, desempeñando así un papel crucial en la regulación del gen P53 (Li *et al.*, 2020).

Otro gen importante en la muestra no tratada (Figuras 5 y 7) es el RPL13A (ENST00000474171.1) que codifica para la proteína ribosomal L13a, que es un componente de la subunidad ribosomal grande 60S. Esta proteína desempeña un papel esencial más allá de la función ribosomal y la síntesis de proteínas, pues forma parte del complejo GAIT (*gamma interferon-activated inhibitor of translation*), un regulador clave en la inhibición de la traducción de transcripciones específicas durante procesos inflamatorios y posee propiedades antitumorales que pueden influir en la progresión de enfermedades como la LLA de

células T (Hou y Tien, 2020). La sobreexpresión de RPL13A y su participación en el complejo GAIT sugieren un mecanismo por el cual las células leucémicas podrían alterar la respuesta inmune y evadir los efectos antiproliferativos del interferón gamma (Herranz, 2022)

En cuanto al modelo animal (Figuras 7 y 9), se visualiza que el transcrito ENST00000651671.1 (NOTCH1) sobresale. Cabe recalcar que NOTCH1 está encargado de la regulación de señales celulares, aumentando la proliferación celular y la aparición de células sanguíneas inmaduras, lo cual promueve la leucemia (Aster, Pear y Blacklow, 2008). La presencia de este gen en esta muestra se debe a que en el diseño del modelo animal se realizó la inserción del oncogén NOTCH1 (NOTCH1-L1601P- Δ PEST) mediante una transducción retroviral de células de médula ósea enriquecidas en células negativas de linaje (Wendorff *et al.*, 2019). Este diseño condujo al desarrollo de LLA en los ratones *knockout*, en los cuales se puede demostrar que la sobreexpresión de NOTCH1 y el silenciamiento del gen PHF6 aumenta la progresión de la enfermedad, así como disminuye la supervivencia en los pacientes, debido a que se sobreexpresan genes que pueden influenciar a la proliferación celular descontrolada como fue el caso del estudio hecho por Yuan *et al.* (2022).

La muestra del modelo animal también cuenta con otro gen sobresaliente de gran importancia, el PTMA (ENST00000409115.8), que codifica para la proteína *prothymosin Alpha*, la cual desempeña un papel en la regulación de procesos celulares como la diferenciación, la apoptosis y la proliferación celular (Ferrara *et al.*, 2010). En la muestra del modelo animal se evidencia una gran cantidad de ARN de este gen, lo cual puede ser a causa de que, como se ha demostrado, PTMA está involucrado en la regulación negativa del proceso apoptótico, lo que puede contribuir a la supervivencia de células malignas, ya que PTMA es un efector aguas debajo de la vía de señalización de c-myc, un proceso que genera el proceso de tumorigénesis (Tanaka y Kusunoki, 2009).

Es de vital importancia continuar con este estudio, especialmente en Ecuador, donde existe un alto número de pacientes con leucemia, en su mayoría niños, que no tienen acceso a trasplantes de médula ósea. La continuidad de esta investigación sería crucial para mejorar la calidad de vida de estos pacientes y desarrollar nuevos enfoques para silenciar los genes involucrados en esta enfermedad.

CONCLUSIONES

La utilización de herramientas como la línea de comandos de bash de Linux es esencial en el estudio de grandes volúmenes de datos, como en el caso de la secuenciación de RNA-seq. Los conjuntos de datos generados son masivos y requieren herramientas eficientes para su procesamiento y análisis. Además, al

momento de realizar un análisis con diferentes especies, el proceso de homologación de las mismas, es un proceso vital, ya que permite tener certeza de que las muestras estudiadas tienen una similitud y que los resultados podrían ser fiables.

Aunque la herramienta Kallisto quant demostró ser esencial para el análisis eficiente de grandes conjuntos de datos en este estudio de RNA-seq, la omisión de una normalización logarítmica plantea desafíos significativos en la interpretación de la expresión diferencial, resaltando la importancia de este proceso para evitar sesgos en los análisis. A pesar de estas dificultades en el proceso de análisis de los datos se pudo visualizar que en las muestras de LLA, existen genes que bibliográficamente han sido asociados con la progresión de esta enfermedad. Por ejemplo, en la muestra tratada se destaca la presencia del gen RPL11, que se encuentra involucrado en la activación de TP53, mientras que en el modelo animal el transcrito de NOTCH1 en el modelo animal, ya que esta es una muestra de LLA de células T, en donde se ha demostrado que este gen se encuentra involucrado en la señalización y diferenciación celular, lo que conlleva a la proliferación de células linfoides.

REFERENCIAS

Aho, A. V., Kernighan, B. W., & Weinberger's, P. J. (1992). The AWK programming language. Includes index (Vol. 9th). Estados Unidos: AT&T.

Andrews, S. (2010). FastQC A quality control tool for high throughput sequence data. Recuperado el 26 de junio de 2023, de Babraham bioinformatics website: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Aster, J. C., Blacklow, S. C., & Pear, W. S. (2011). Notch signalling in T-cell lymphoblastic leukaemia/lymphoma and other haematological malignancies. *The Journal of pathology*, 223(2), 262–273. <https://doi.org/10.1002/path.2789>

Aster, J. C., Pear, W. S., & Blacklow, S. C. (2008). Notch signaling in leukemia. *Annual review of pathology*, 3, 587–613. <https://doi.org/10.1146/annurev.pathmechdis.3.121806.154300>

Autry, R. J., Paugh, S. W., Carter, R., Shi, L., Liu, J., Ferguson, D. C., Lau, C. E., Bonten, E. J., Yang, W., McCorkle, J. R., Beard, J. A., Panetta, J. C., Diedrich, J. D., Crews, K. R., Pei, D., Coke, C. J., Natarajan, S., Khatamian, A., Karol, S. E., Lopez-Lopez, E., ... Evans, W. E. (2020). Integrative genomic analyses reveal mechanisms of glucocorticoid resistance in acute lymphoblastic leukemia. *Nature cancer*, 1(3), 329–344. <https://doi.org/10.1038/s43018-020-0037-3>

Barz, M. J., Behrmann, L., Capron, D., Zuchtriegel, G., Steffen, F. D., Kunz, L., Zhang, Y., Vermeerbergen, I. J., Marovca, B., Kirschmann, M., Zech, A., Nombela-Arrieta, C., Ziegler, U., Schroeder, T., Bornhauser, B., & Bourquin, J. P. (2023). B- and T-cell acute lymphoblastic leukemias evade chemotherapy at distinct sites in the bone marrow. *Haematologica*, 108(5), 1244–1258. <https://doi.org/10.3324/haematol.2021.280451>

- Blankenberg, D., Coraor, N., Von Kuster, G., Taylor, J., Nekrutenko, A., & The Galaxy Team. (2011). Integrating diverse databases into a unified analysis framework: a Galaxy approach. Database, 2011, bar011. <https://doi.org/10.1093/database/bar011>
- Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible read trimming tool for Illumina NGS data. Recuperado el 26 de junio de 2023, de Usadellab.org website: <http://www.usadellab.org/cms/?page=trimmomatic>
- Bray, N., Pimentel, H., Melsted, P. et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527 (2016). <https://doi.org/10.1038/nbt.3519>
- Breschi A, Gingeras TR, Guigó R. Comparative transcriptomics in human and mouse. Nat Rev Genet. 2017 Jul;18(7):425-440. doi: 10.1038/nrg.2017.19. Epub 2017 May 8. PMID: 28479595; PMCID: PMC6413734.
- Chennamadhavuni, A., Lyengar, V., Mukkamalla, S. K. R., & Shimanovsky, A. (2023). Leukemia. In StatPearls. StatPearls Publishing.
- CHU Sainte-Justine (2016). The genomic landscape of pediatric T-cell acute lymphoblastic leukemia reveals novel X-linked somatic mutations associated with apoptosis resistance. [RNA-Seq], geo, V1. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78783>.
- Dai, M.-S., Shi, D., Jin, Y., Sun, X.-X., Zhang, Y., Grossman, S. R., & Lu, H. (2006). Regulation of the MDM2-p53 pathway by ribosomal protein L11 involves a post-ubiquitination mechanism. The Journal of Biological Chemistry, 281(34), 24304–24313. doi:10.1074/jbc.M602596200
- Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research, 30(1), 207–210. <https://doi.org/10.1093/nar/30.1.207>
- Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Briefings in bioinformatics, 19(5), 776–792. <https://doi.org/10.1093/bib/bbx008>
- Ferrara, D., Izzo, G., Pariante, P., Donizetti, A., d'Istria, M., Aniello, F., & Minucci, S. (2010). Expression of prothymosin alpha in meiotic and post-meiotic germ cells during the first wave of rat spermatogenesis. Journal of cellular physiology, 224(2), 362–368. <https://doi.org/10.1002/jcp.22131>
- Grimwade, D., Ivey, A., & Huntly, B. J. (2016). Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood, 127(1), 29–41. <https://doi.org/10.1182/blood-2015-07-604496>
- Gutiérrez, A., & Kentsis, A. (2018). Acute myeloid/T-lymphoblastic leukaemia (AMTL): a distinct category of acute leukaemias with common pathogenesis in need of improved therapy. British Journal of Haematology, 180(6), 919–924. doi:10.1111/bjh.15129
- Herranz D. (2022). INTERFERing with the progression of T-cell acute lymphoblastic leukemia: a multifaceted therapy. Haematologica, 107(6), 1232–1234. <https://doi.org/10.3324/haematol.2021.279549>

Hou, H. A., & Tien, H. F. (2020). Genomic landscape in acute myeloid leukemia and its implications in risk classification and targeted therapies. *Journal of Biomedical Science*, 27, 81. <https://doi.org/10.1186/s12929-020-00674-7>

Inaba, H., & Pui, C.-H. (2010). Glucocorticoid use in acute lymphoblastic leukaemia. *The Lancet Oncology*, 11(11), 1096–1106. doi:10.1016/S1470-2045(10)70114-5

Jiang, L., Yuan, C. M., Hubacheck, J., Janik, J. E., Wilson, W., Morris, J. C., Jasper, G. A., & Stetler-Stevenson, M. (2009). Variable CD52 expression in mature T cell and NK cell malignancies: implications for alemtuzumab therapy. *British journal of haematology*, 145(2), 173–179. <https://doi.org/10.1111/j.1365-2141.2009.07606.x>

Juliusson, G., & Hough, R. (2016). Leukemia. *Progress in tumor research*, 43, 87–100. <https://doi.org/10.1159/000447076>

Kim, K. M., Ahn, A. R., Park, H. S., et al. (2022). *Clinical significance of p53 protein expression and TP53 variation status in colorectal cancer. BMC Cancer*, 22, 940.

Liang, D. C., Shih, L. Y., Fu, J. F., Li, H. Y., Wang, H. I., Hung, I. J., Yang, C. P., Jaing, T. H., Chen, S. H., & Liu, H. C. (2006). K-Ras mutations and N-Ras mutations in childhood acute leukemias with or without mixed-lineage leukemia gene rearrangements. *Cancer*, 106(4), 950–956. <https://doi.org/10.1002/cncr.21687>

Li, Y., Chen, F., Shen, W., Li, B., Xiang, R., Qu, L., Zhang, C., Li, G., Xie, H., Katanaev, V. L., & Jia, L. (2020). WDR74 induces nuclear β -catenin accumulation and activates Wnt-responsive genes to promote lung cancer growth and metastasis. *Cancer Letters*, 471, 103–115. <https://doi.org/10.1016/j.canlet.2019.12.011>

Li, Y., Yang, X., Sun, Y., Li, Z., Yang, W., Ju, B., Easton, J., Pei, D., Cheng, C., Lee, S., Pui, C. H., Yu, J., Chi, H., & Yang, J. J. (2022). Impact of T-cell immunity on chemotherapy response in childhood acute lymphoblastic leukemia. *Blood*, 140(13), 1507–1521. <https://doi.org/10.1182/blood.2021014495>

Li, Y., Zhou, Y., Li, B., Chen, F., Shen, W., Lu, Y., Zhong, C., Zhang, C., Xie, H., Katanaev, V. L., & Jia, L. (2020). WDR74 modulates melanoma tumorigenesis and metastasis through the RPL5-MDM2-p53 pathway. *Oncogene*, 39(13), 2741–2755. <https://doi.org/10.1038/s41388-020-1179-6>

Liu, Y., Deisenroth, C., & Zhang, Y. (2016). RP-MDM2-p53 pathway: Linking ribosomal biogenesis and tumor surveillance. *Trends in Cancer*, 2(4), 191–204. doi:10.1016/j.trecan.2016.03.002

Liu J, Zhao M, Yuan B, Gu S, Zheng M, Zou J, Jin J, Liu T, Feng XH. WDR74 functions as a novel coactivator in TGF- β signaling. *J Genet Genomics*. 2018 Dec 20;45(12):639-650. doi: 10.1016/j.jgg.2018.08.005. Epub 2018 Nov 27. PMID: 30594465.

Medina, D., Castillo, A., Beltrán, E., Manzi, E., Chantre-Mostacilla, A., Guerrero-Fajardo, G. P., ... Franco, A. A. (2022). Impacto de la calidad de vida relacionada a la salud en pacientes pediátricos con trasplante de progenitores hematopoyéticos en una institución colombiana. *Biomedica: Revista Del Instituto Nacional de Salud*, 42(3), 531–540. doi:10.7705/biomedica.6403

Ministerio de Salud Pública del Ecuador. (2023). Crónica de una guerrera, de 11 años, que enfrenta el cáncer – Ministerio de Salud Pública. Recuperado el 6 de septiembre de 2023, de Gob.ec website: <https://www.salud.gob.ec/cronica-de-una-guerrera-de-11-anos-que-enfrenta-el-cancer/>

Mohanty, S., & Heuser, M. (2021). Mouse Models of Frequently Mutated Genes in Acute Myeloid Leukemia. *Cancers*, 13(24), 6192. <https://doi.org/10.3390/cancers13246192>

Mora-Márquez, F., Vázquez-Poletti, J. L., & López de Heredia, U. (2021). NGScloud2: optimized bioinformatic analysis using Amazon Web Services. *PeerJ*, 9(e11237), e11237. doi:10.7717/peerj.11237

Morgulis A., Coulouris G., Raytselis Y., Madden T.L., Agarwala R., Schaffer A.A. (2008) "Database indexing for production MegaBLAST searches." *Bioinformatics* 15:1757-1764.

Prada-Arismendy, J., Arroyave, J. C., & Röthlisberger, S. (2017). Molecular biomarkers in acute myeloid leukemia. *Blood reviews*, 31(1), 63–76. <https://doi.org/10.1016/j.blre.2016.08.005>

National Center for Biotechnology Information, U.S. National Library of Medicine. (2024, Marzo 5). CD52 molecule [Homo sapiens (human)]. NCBI. Recuperado el 12 de marzo de 2024, de <https://www.ncbi.nlm.nih.gov/gene/1043>

NIH Genomics. (2022). Downloading SRA toolkit. Recuperado el 26 de junio de 2023, de GitHub website: <https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. doi:10.1093/bioinformatics/btp616

Sekeres, M. A., Guyatt, G., Abel, G., Alibhai, S., Altman, J. K., Buckstein, R., Choe, H., Desai, P., Erba, H., Hourigan, C. S., LeBlanc, T. W., Litzow, M., MacEachern, J., Michaelis, L. C., Mukherjee, S., O'Dwyer, K., Rosko, A., Stone, R., Agarwal, A., Colunga-Lozano, L. E., ... Brignardello-Petersen, R. (2020). American

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: a cancer journal for clinicians*, 73(1), 17–48. <https://doi.org/10.3322/caac.21763>

Shah, D. S., & Kumar, R. (2013). Steroid resistance in leukemia. *World journal of experimental medicine*, 3(2), 21–25. <https://doi.org/10.5493/wjem.v3.i2.21>

Stallman, R. M., & Granlund, T. (1997). split(1): split file into pieces - Linux man page. Recuperado el 22 de julio de 2023, de Die.net website: <https://linux.die.net/man/1/split>

Tanaka, K., & Kusunoki, M. (2009). PTMA (prothymosin, alpha). *Atlas Genet Cytogenet Oncol Haematol*. Recuperado el 12 de marzo de 2024, de <http://atlasgeneticsoncology.org/gene/44094/>

Teachey, D. T., & Pui, C. H. (2019). Comparative features and outcomes between paediatric T-cell and B-cell acute lymphoblastic leukaemia. *The Lancet. Oncology*, 20(3), e142–e154. [https://doi.org/10.1016/S1470-2045\(19\)30031-2](https://doi.org/10.1016/S1470-2045(19)30031-2)

Vlachos A. (2017). Acquired ribosomopathies in leukemia and solid tumors. *Hematology. American Society of Hematology. Education Program*, 2017(1), 716–719. <https://doi.org/10.1182/asheducation-2017.1.716>

Wendorff, A. A., Quinn, S. A., Rashkovan, M., Madubata, C. J., Ambesi-Impiombato, A., Litzow, M. R., ... Ferrando, A. A. (2018). Phf6 loss enhances HSC self-renewal driving tumor initiation and leukemia stem cell activity in T-ALL. *Cancer Discovery*, 9(3), 436–451. doi:10.1158/2159-8290.cd-18-1005

Xiong, X., Liu, X., Li, H. et al. (2018). Ribosomal protein S27-like regulates autophagy via the β -TrCP-DEPTOR-mTORC1 axis. *Cell Death & Disease*, 9, 1131. <https://doi.org/10.1038/s41419-018-1168-7>

Xu, X., Xiong, X., & Sun, Y. (2016). The role of ribosomal proteins in the regulation of cell proliferation, tumorigenesis, and genomic integrity. *Science China. Life sciences*, 59(7), 656–672. <https://doi.org/10.1007/s11427-016-0018-0>

Yuan, S., Wang, X., Hou, S. et al. (2022). PHF6 and JAK3 mutations cooperate to drive T-cell acute lymphoblastic leukemia progression. *Leukemia* 36, 370–382 [https://doi.org/10.1038/s41375-021-01392-](https://doi.org/10.1038/s41375-021-01392-1)