

Identificación de los polimorfismos de inserción de elementos transponibles (TIPs) del retrovirus endógeno humano K (HERV-K) y su relación con leucemia mieloide aguda (LMA), por medio de análisis *in silico*.

Nicolás Camargo Forero

Trabajo de Grado para Optar el Título de Biólogo

Director

Simón Orozco Arias

Ingeniero de Sistemas

Codirector

Romain Guyot

Doctor en Genética

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2021

Dedicatoria

A mi madre, Adriana Forero Meza, quien me apoyó constantemente día a día, liberándome de cargas para facilitar mi vida personal y universitaria, dándome fuerzas para continuar y salir adelante.

A mi padre, Gerardo Camargo Contreras, quien con su constancia, perseverancia y fuerza mental me enseñó cómo mejorar emocional y profesionalmente.

A mi hermano, Leonardo Camargo Forero, mi ejemplo a seguir y mi segundo padre, quien desde niño me enseñó tantas cosas que me hacen ser la persona que soy.

A Carolina Torres, quien me acompañó y apoyó desde el inicio hasta el final de la carrera.

A mis amigos, quienes hicieron más llevaderos los momentos difíciles y más amenos los momentos alegres.

Agradecimientos

Al Instituto de Investigación para el Desarrollo (IRD) de Francia por facilitarme el acceso al supercomputador Itrop y sus recursos.

A mi tutor Simón Orozco, quien me apoyo en el desarrollo de todo el proyecto.

A mi tutor Romain Guyot por brindar su experiencia en las áreas de la biología y la informática.

A mi hermano Leonardo, quien me apoyó en los momentos en los que estuve estancado, brindándome su conocimiento.

A mi familia, quienes me apoyaron incondicionalmente.

Tabla de Contenido

	Pág.
Introducción	10
1. Objetivos	13
1.1 Objetivo General.....	13
1.2 Objetivos Específicos.....	13
2. Cuerpo del trabajo.....	14
2.1 Marco Referencial.....	14
2.1.1 Método	25
2.1.1.1 Descripción del conjunto de datos.	25
2.1.1.2 Identificación de los TIPS de HERV-K	27
2.1.1.3 Análisis estadístico.....	29
2.1.1.4 Identificación de posibles genes afectados por los TIPS.....	29
2.1.2 Resultados	30
2.1.2.1 Discusión.....	47
3. Conclusiones.....	56
Referencias bibliográficas.....	58

Lista de Tablas

	Pág.
Tabla 1. Pipeline de TIP_finder	27
Tabla 2. TIPS de HERV-K asociados a leucemia mieloide aguda	31
Tabla 3. Características de los genes encontrados en la región de interés	35

Lista de Figuras

	Pág.
Figura 1. Clases de elementos transponibles, estructura y porcentaje del genoma humano	15
Figura 2. Estructura de un provirus de HERV-K	18
Figura 3. Polimorfismo de inserción de elemento transponible (TIP)	22
Figura 4. Lecturas discordantes	28
Figura 5. TIPs de HERV-K en el genoma completo de pacientes de LMA y controles	32
Figura 6. TIPs de HERV-K asociados a LMA en el cromosoma 8 de pacientes de LMA y controles.....	33
Figura 7. Navegador de genoma para la región de interés del cromosoma 8	34
Figura 8. Navegador de genoma para el gen codificante <i>TRIB1</i>	39
Figura 9. Navegador de genoma para el gen codificante <i>LRATD2</i>	40
Figura 10. Navegador de genoma para el gen codificante <i>POU5F1B</i>	41
Figura 11. Navegador de genoma para el gen codificante <i>MYC</i>	42
Figura 12. Navegador de genoma para el gen no codificante <i>PCATI</i>	43
Figura 13. Navegador de genoma para el gen no codificante <i>PVT1</i>	44
Figura 14. Navegador de genoma para el gen no codificante <i>CCDC26</i>	45
Figura 15. Metodología para la identificación de TIPs de una familia de elementos transponibles y su correlación con enfermedades específicas	46
Figura 16. Cantidad de TIPs de HERV-K asociados a LMA por condición	54

Lista de Apéndices

Los apéndices están adjuntos y puede visualizarlos en la base de datos de la biblioteca

UIS

Apéndice A. Lista de pacientes de leucemia mieloide aguda

Apéndice B. Lista de controles

Apéndice C. Librería de HERV-K

Apéndice D. Matriz final de TIPS para pacientes de leucemia mieloide aguda

Apéndice E. Matriz final de TIPS para controles

Apéndice F. TIPS de HERV-K asociados a leucemia mieloide aguda

Apéndice G. Navegador de genoma para el gen no codificante *CASC8*

Apéndice H. Navegador de genoma para el gen no codificante *LINC00824*

Apéndice I. Navegador de genoma para el gen no codificante *PNCRI*

Resumen

Título: Identificación de los polimorfismos de inserción de elementos transponibles (TIPs) del retrovirus endógeno humano K (HERV-K) y su relación con leucemia mieloide aguda (LMA), por medio de análisis *in silico**

Autor: Nicolás Camargo Forero**

Palabras Clave: Elementos transponibles, retrovirus endógenos humanos, HERV-K, leucemia mieloide aguda, TIPs.

Descripción: Los retrovirus endógenos humanos (HERV) son retrotransposones LTR que se insertaron en el genoma humano mediante la infección viral exógena. A través de mutaciones, replicación y herencia mendeliana, dichos retrovirus componen actualmente el 8% del genoma humano. La familia HERV-K puede tener hasta cuatro marcos abiertos de lectura (ORF) que codifican las proteínas Gag, Pro, Pol, Env/Rec/Np9, las cuales se han visto implicadas en la carcinogénesis de diferentes tipos de cáncer tales como: cáncer de ovario, próstata, mama y leucemia. Dentro de los tipos de leucemia cabe resaltar la Leucemia Mieloide Aguda (LMA), la cual es un problema a nivel mundial debido a que representa el 1% de las muertes por cáncer en Estados Unidos y tiene una tasa de supervivencia del 29,5%. Por esta razón, se presenta una metodología para identificar la asociación estadística entre polimorfismos de inserción de elementos transponibles (TIPs) de HERV-K y leucemia mieloide aguda usando lecturas de secuenciación de genomas completos de pacientes y el software TIP_finder. Se encontraron 101 TIPs de HERV-K asociados a la LMA en un porcentaje entre el 44,4 al 56,6%, los cuales en su mayoría (70) se encontraban ubicados en la región 8q24.13 y 8q24.21. Además, en promedio cada paciente de LMA presentaba 80 TIPs de HERV-K asociados mientras que en los controles no había presencia de estos. Por otro lado, se encontró que los genes *TRIB1*, *LRATD2*, *POU5F1B*, *MYC*, *PCAT1*, *PVT1* y *CCDC26* podrían verse afectados por estas inserciones al causar desplazamiento o fragmentación de sus regiones de interés. Además, se propone una metodología para analizar la asociación entre una familia de elementos transponibles y una enfermedad específica. Finalmente, a través del análisis estadístico de X^2 ($\alpha=0.05$) se identificó que los TIPs de HERV-K reportados en este estudio están asociados estadísticamente con la leucemia mieloide aguda.

*Trabajo de grado

**Facultad de Ciencias. Escuela de Biología. Director: Simón Orozco Arias. Ingeniero de Sistemas. Codirector: Romain Guyot. Doctor en genética

Abstract

Title: Identification of transposable element insertion polymorphisms (TIPs) of human endogenous retrovirus K (HERV-K) and its correlation to acute myeloid leukemia (AML) by *in silico* analysis.*

Author: Nicolás Camargo Forero**

Key Words: Transposable elements, human endogenous retrovirus, HERV-K, acute myeloid leukemia, TIPs.

Description: Human Endogenous RetroVirus (HERV) are Long Terminal Repeats (LTR) retrotransposons that were inserted into the human genome through exogenous viral infection. Through mutation, replication and Mendelian inheritance, such retroviruses currently make up to 8% of the human genome. The HERV-K family can have up to four Open Reading Frames (ORF) encoding Gag, Pro, Pol, Env/Rec/Np9 proteins, which have been implicated in the carcinogenesis of different types of cancer such as ovarian, prostate, breast and leukemia, among others. Considering Leukemia, it is worth highlighting Acute Myeloid Leukemia (AML), a considerable problem given that it represents 1% of cancer deaths in the United States and has a survival rate of 28.7%. In this work, we present a methodology to identify the statistical association between HERV-K Transposable element Insertion Polymorphisms (TIPs) and acute myeloid leukemia using whole genome sequencing reads (patients and controls) via the TIP_finder software. Our results show that 101 HERV-K TIPs are associated with AML with a percentage between 44.4 to 56.6%, most of which (70) were located in the region from 8q24.13 to 8q24.21. In addition, it was found that on average each AML patient had 80 HERV-K associated TIPs which were not found in control patients. Moreover, it was found that *TRIB1*, *LRATD2*, *POU5F1B*, *MYC*, *PCAT1*, *PVT1* and *CCDC26* genes and its regions of interest could be displaced or fragmented by TIPs insertions. Furthermore, a general methodology was devised to facilitate the association analysis between transposable elements families and specific diseases. Finally, we conclude that the HERV-K TIPs reported in this study are statistically correlated with acute myeloid leukemia as confirmed by the X^2 test with alpha equals 0.05.

*Bachelor thesis

** Science faculty. School of biology. Director: Simón Orozco Arias. Systems engineer. Co-director: Romain Guyot. PhD. in genetics.

Introducción

La leucemia mieloide aguda (LMA), es un tipo de cáncer producido por la rápida proliferación de células anormales de la línea celular mieloide de origen clonal que infiltran la médula ósea, sangre periférica y otros tejidos, en donde su acumulación interfiere con la producción de glóbulos rojos normales (Leyto-Cruz, 2018). Se ha reportado que la LMA es el tipo de leucemia más común en adultos, comprendiendo el 1,1% de todos los casos nuevos de cáncer (Howlader et al., 2020; Redaelli et al., 2003), y el 1,8% de las muertes por cáncer en Estados Unidos (Howlader et al., 2020; Leyto-Cruz, 2018), asimismo, un estudio realizado por el Programa de Vigilancia, Epidemiología y Resultados Finales (SEER, por sus siglas en inglés) estimó que en el 2020 se presentaron cerca de 20.000 nuevos casos de leucemia mieloide aguda en Estados Unidos (Howlader et al., 2020). Además, un estudio epidemiológico realizado en 2015 por el Instituto Nacional de Cancerología, reportó que la LMA representa el 10% de los casos de cáncer en personas menores de 18 años en Colombia (Reyes et al., 2016).

Por otro lado, los retrovirus endógenos humanos (*Human Endogenous RetroVirus*), son elementos transponibles insertados en el genoma humano mediante la infección viral exógena (Subramanian et al., 2011) y en su mayoría se encuentran mutados o incompletos, lo cual les impide codificar genes funcionales (García-Montojo et al., 2018). No obstante, la familia HERV-K puede tener hasta cuatro marcos abiertos de lectura que codifican las proteínas Gag, Pro, Pol, Env/Rec/Np9 (Cegolon et al., 2013). Numerosos estudios han identificado que la expresión de los genes de retrovirus endógenos humanos de tipo K (HERV-K), se encuentran asociados con diferentes tipos de leucemia, tales como: leucemia mieloide crónica (LMC) (Brodsky et al., 1993), leucemia linfocítica crónica (LLC) (Depil et al., 2002), leucemia linfoblástica aguda (LLA)

(Bergallo et al., 2017; Januszkiewicz-Lewandowska et al., 2013) y leucemia mieloide aguda (LMA) (Depil et al., 2002; Januszkiewicz-Lewandowska et al., 2013).

Los HERV, al igual que todos los elementos transponibles, son muy diversos y se encuentran en diferentes partes del genoma, por lo que se han realizado estudios para identificar los polimorfismos de inserción de elementos transponibles (TIPs, por sus siglas en inglés) en el genoma completo (Bradley-Golem, 2013; X. Chen et al., 2019). Los TIPs son inserciones de elementos transponibles en el genoma de un individuo, en una posición que no se presenta en el genoma de referencia, los cuales se han visto asociados con diferentes fenómenos biológicos, incluyendo adaptación, mutación, silenciamiento de genes y regulación de la expresión génica (Barrón et al., 2014; Carpentier et al., 2019; Drongitis et al., 2019).

Sin embargo, en la literatura, los análisis realizados de asociación de HERV-K y leucemia son basados en la sobre expresión, desregulación o sub expresión de alguno de sus genes (Bergallo et al., 2017; Brodsky et al., 1993; Depil et al., 2002; Januszkiewicz-Lewandowska et al., 2013) y no en la asociación de los TIPs de estos en el genoma completo con los diferentes tipos de cáncer, entre estos, la leucemia mieloide aguda.

Los avances tecnológicos y computacionales han permitido que el análisis *in silico* resulte más eficiente en términos de tiempo y costo económico en comparación con análisis *in vivo* o *in vitro* (Mangul et al., 2019). En este sentido, se han utilizado herramientas bioinformáticas para analizar datos genómicos de pacientes, lo que ha permitido generar nuevas estrategias de terapias génicas para diferentes tipos de cáncer (Ding et al., 2017; Lu et al., 2018; Xue et al., 2020), incluyendo la leucemia mieloide aguda (Feng et al., 2018; Wei Li et al., 2017). La creación de nuevas estrategias para terapias génicas contra el cáncer, el diagnóstico temprano o predicción del mismo basado en genes, SNPs, o incluso genomas completos, son algunos

objetivos de la unión entre la biología y la informática (bioinformática) y su aporte en la medicina (Chande et al., 2018; Y. Liu et al., 2019; Rong et al., 2018; Shao et al., 2019).

Actualmente, hay varios softwares usados para la identificación de TIPS en el genoma completo de un organismo, los cuales, en su mayoría, están basados en la localización discordante de lecturas pareadas y divididas (Ewing, 2015; Orozco-Arias et al., 2020), entre estos TRAKCPOSON, el cual permite identificar TIPS en conjuntos de datos grandes tales como el genoma del arroz (400 Mb), sin embargo, este software no es eficiente al momento de analizar genomas de mayor tamaño como el del humano con 3.3 Gb. Por esto, Orozco-arias et al., (2020) presentaron el software TIP_finder, que utiliza la misma metodología del software TRACKPOSON e implementa, a su vez, técnicas de computación de alto rendimiento y programación en paralelo, permitiendo así, el análisis eficiente de genomas de gran tamaño, lo que lo hizo óptimo para realizar este estudio.

Por esta razón, identificar relaciones entre TIPS de HERV-K y leucemia mieloide aguda, podría generar una nueva estrategia para diagnóstico y predicción de cáncer basándose en el genoma completo. Además, la metodología utilizada puede replicarse para diferentes tipos de cáncer y otras enfermedades asociadas a las familias de HERV u otros elementos transponibles. De esta manera, el objetivo de este estudio consiste en evaluar la asociación entre polimorfismos de inserción de elementos transponibles (TIPS) del retrovirus endógeno humano K (HERV-K) y leucemia mieloide aguda (LMA), por medio de análisis *in silico*, y así, contribuir al diseño de nuevas estrategias de diagnóstico temprano y predicción de esta enfermedad.

1. Objetivos

1.1 Objetivo General

Evaluar la asociación entre polimorfismos de inserción de elementos transponibles (TIPs) del retrovirus endógeno humano K (HERV-K) y leucemia mieloide aguda (LMA).

1.2 Objetivos Específicos

Identificar los TIPs de HERV-K asociados a leucemia mieloide aguda mediante el estudio de genomas de pacientes.

Calcular el grado de asociación entre cada TIP de HERV-K encontrado y la leucemia mieloide aguda.

Examinar genes potencialmente afectados por la existencia de los TIPs relacionados a leucemia mieloide aguda.

2. Cuerpo del Trabajo

2.1 Marco Referencial

Los elementos transponibles (TE, por sus siglas en inglés) pueden definirse como una secuencia genómica que tiene la capacidad de cambiar su localización (transponerse) dentro del genoma (Wen-hsiung Li, 1997), sin embargo, esta definición no considera dos hechos de los elementos transponibles, los cuales son: no todos los elementos transponibles tienen la capacidad de cambiar su localización por sí mismos y requieren enzimas codificadas de otros elementos para realizarlo y, la mayoría de elementos no cambian su localización, sino que son copiados y las copias son las que se insertan en diferentes posiciones. Por lo tanto, una definición más acertada de elementos transponibles es una secuencia genómica que puede separarse de su ubicación original en el genoma y reinsertarse en otra, o que puede copiarse e insertar sus copias en otros lugares del genoma, ya sea autónomamente o empleando enzimas de otros elementos transponibles (Tubio, 2009).

Los TE pueden encontrarse en casi todas las especies eucariotas conocidas en la actualidad (Wicker et al., 2007), y su papel es clave en la diversificación del genoma y la adaptación y evolución de los organismos, al generar mutaciones y amplificar el número de sus copias (Casacuberta & González, 2013; Todorovska, 2007). Además, los TE son muy diversos, con más de 1000 familias que componen una gran parte del genoma de los eucariotas, constituyendo más del 80% del genoma de las plantas, del 3 al 20% del genoma de los hongos y del 3 al 45% del genoma de los metazoos (Wicker et al., 2007). Finalmente, los elementos transponibles pueden clasificarse en clases, dependiendo del mecanismo de transposición, siendo clase I los elementos que se transponen usando una molécula de ARN y clase II, los elementos

que se transponen usando una molécula de ADN (Figura 1) (Casacuberta & González, 2013; Wicker et al., 2007).

Figura 1

Clases de elementos transponibles, estructura y porcentaje del genoma humano

Clase			% del genoma humano
I	Retrotransposones LTR (ERVs)		8
I	Retrotransposones no LTR (LINE-1)		17
II	Transposones		3

Los elementos transponibles clase I, también llamados retrotransposones, se caracterizan por la transcripción de un ARN intermediario a partir de una copia genómica, el cual realiza retrotranscripción o transcripción inversa mediante retrotranscriptasa para insertar esta copia en el genoma, siendo así, los principales contribuyentes a la parte repetitiva del genoma (Wicker et al., 2007). Los TE clase I pueden clasificarse en cinco órdenes: elementos similares a DIRS (*Dictyostelium Intermediate Repeat Sequence*), elementos similares a Penelope (PLEs, por sus siglas en inglés), elementos largos intercalados (LINEs, por sus siglas en inglés), elementos cortos intercalados (SINEs, por sus siglas en inglés) y retrotransposones LTR (Makałowski et al., 2019).

Estos últimos se caracterizan por tener repeticiones terminales largas (LTR, por sus siglas en inglés), que son secuencias de ADN no codificantes, localizadas en ambas terminales de la región codificante interna (Orozco-Arias et al., 2019). Además, al igual que los retrovirus exógenos, contienen un gen *Gag* que codifica una capa de partículas virales y un gen *Pol* que codifica la transcriptasa inversa, la ribonucleasa H y la integrasa, que son las enzimas encargadas

de la retrotranscripción y la integración de la copia al genoma. Sin embargo, a diferencia de los retrovirus exógenos, la mayoría de los retrotransposones LTR no contienen un gen *Env* (al menos no funcional) que codifica la envoltura y facilita la migración a otras células. Los retrotransposones LTR pueden clasificarse en cinco superfamilias, las cuales son: Copia, Gypsy, Bel-Pao, retrovirus y retrovirus endógenos (ERV, por sus siglas en inglés) (Makałowski et al., 2019).

Los retrovirus endógenos humanos (HERV, por sus siglas en inglés), son retrotransposones LTR que se insertaron en el genoma humano mediante la infección viral exógena en las células germinales de los primates hace millones de años (Hohn et al., 2013; Subramanian et al., 2011). Actualmente componen el 8% del genoma humano, con más de 500.000 elementos transponibles debido a mutaciones, replicación y herencia mendeliana (Garcia-Montojo et al., 2018). Sin embargo, en su mayoría se encuentran mutados o incompletos, inhabilitándoles para codificar genes funcionales. Por esta razón, los HERV pueden encontrarse en forma proviral de longitud completa o casi completa o como LTRs solitarios, que son resultado de la recombinación entre los dos LTR que rodean el provirus y la eliminación de la secuencia interna (Hughes & Coffin, 2004). Además, los ERVs pueden clasificarse en ERV1, los cuales son Gammaretrovirus similares a los virus de la leucemia murina (MLV, por sus siglas en inglés); ERV2, los cuales son Betaretrovirus similares al virus del tumor mamario del ratón (MMTV, por sus siglas en inglés); ERV3 o elementos similares a Spumaretrovirus como el virus espumoso de los simios; Lentivirus; y ERV4, que no tienen ninguna contraparte exógena descrita (Garcia-Montojo et al., 2018).

Entre los retrovirus endógenos ERV2, cabe destacar la familia HERV-K, la cual es la más recientemente adquirida y biológicamente activa, conteniendo hasta cuatro marcos abiertos de

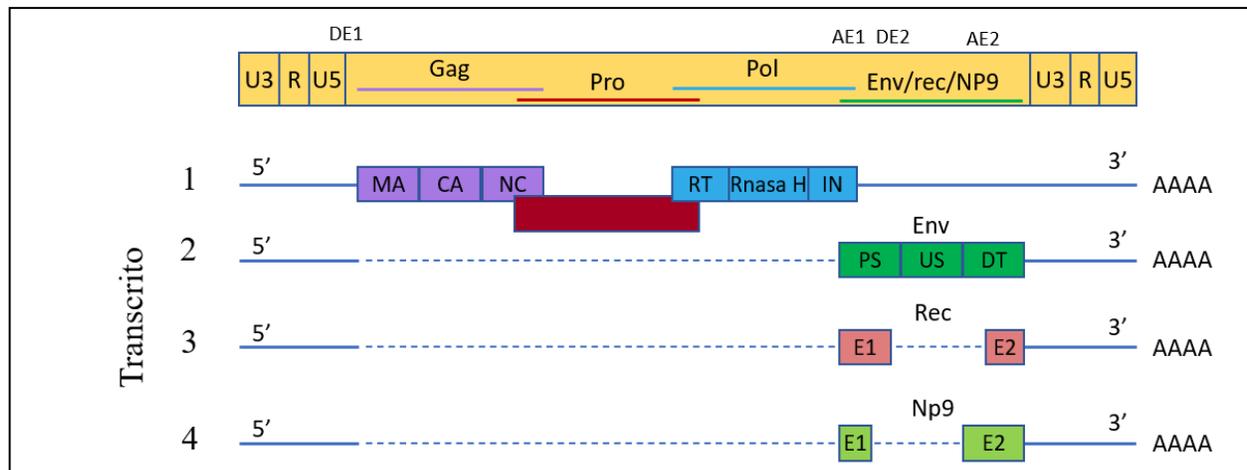
lectura (ORF, por sus siglas en inglés) que codifican las proteínas Gag, Pro, Pol, Env/Rec/Np9 (Cegolon et al., 2013; Garcia-Montojo et al., 2018). La denominación K proviene del uso de un ARNt de lisina (K, según el código de una letra IUPAC) para el cebado de la retrotranscripción (Hohn et al., 2013). A su vez, la familia HERV-K puede dividirse en 10 grupos denominados HML-X (HML = *Human Mouse Mammary tumor virus Like*), donde la X especifica un número entre 1 y 10, y su nomenclatura se basó en la similitud existente entre el grupo y el virus del tumor mamario del ratón (MMTV) (Garcia-Montojo et al., 2018; Hohn et al., 2013). Actualmente, el grupo más relevante de la familia es el HML-2, debido a sus posibles funciones fisiológicas como: el mantenimiento de la pluripotencia, efectos neurotróficos, defensa antiviral y la formación de la placenta; además, de su relación con múltiples enfermedades, incluyendo diferentes tipos de cáncer, esquizofrenia, esclerosis lateral amiotrófica, VIH, entre otras (Garcia-Montojo et al., 2018).

El grupo HML-2, al igual que todos los retrovirus endógenos puede encontrarse en las formas de provirus y LTR solitario, en donde las estimaciones más recientes sugieren que hay 89 provirus de HML-2 y entre 944 y 1.200 LTR solitarios (Garcia-Montojo et al., 2018). Por otra parte, los provirus de HML-2 pueden dividirse en dos tipos dependiendo de la presencia o ausencia de una delección de 292 pares de bases en la unión de los genes *Pol-Env*. Por consiguiente, los provirus de tipo 1 (26% del total de los provirus HML-2) que poseen la delección, son incapaces de codificar Rec o Env debido a la pérdida de un donador de empalme (*splicing donor*, en inglés), no obstante, debido a esta delección se utiliza otro donador de empalme que codifica para la proteína Np9. Finalmente, los provirus de tipo 2 (74% del total de los provirus HML-2) que no poseen la delección, pueden codificar las proteínas Env y Rec, la cual

es una proteína que se une a los transcritos virales facilitando su transporte del núcleo al citoplasma (García-Montojo et al., 2018; Hohn et al., 2013).

Figura 2

Estructura de un provirus de HERV-K



Como se observa en la figura 2, un provirus de HML-2 está compuesto por dos LTRs rodeando la secuencia de los cuatro marcos abiertos de lectura, dos sitios donadores de empalme (DE) y dos sitios aceptores de empalme (AE). Cada LTR consiste en tres regiones (U3, R y U5) en dirección 5' a 3', que poseen elementos promotores y potenciadores de la transcripción, sitios de unión del factor de transcripción y señales de poliadenilación (Hohn et al., 2013). Por otro lado, dependiendo del tipo de provirus se pueden presentar cuatro transcritos, en donde el primero, tiene 3 marcos abiertos de lectura que codifican para la poliproteína Gag-Pro-Pol, en donde *Gag* contiene la secuencia para formar las proteínas de la matriz (MA), cápside (CA) y nucleocápside (NC), *Pro* codifica para la enzima dUTPasa y *Pol* contiene las proteínas de retrotranscriptasa (RT), Rnasa H e integrasa (IN); el segundo, codifica la proteína Env que se compone de tres dominios: el péptido señal (PS), la unidad de superficie (US) y el dominio transmembrana (DT); el tercero, es el producto de un empalme alternativo (*alternative splicing*)

del marco abierto de lectura del gen *Env* que codifica para la proteína Rec, compuesta de dos exones (E), el cual solo es producido por los provirus tipo 2; y el último, debido a la delección de los provirus tipo 1, se pierde el segundo sitio donador de empalme, por lo que se utiliza un sitio alternativo que codifica la proteína Np9 (Garcia-Montojo et al., 2018).

Como se mencionó anteriormente, el grupo HML-2 ha sido de controversia debido a su relación con varias enfermedades, incluyendo diferentes tipos de cáncer. La sobreexpresión o desregulación del gen *Gag* se ha visto asociada con cáncer de próstata (Reis et al., 2013; Wallace et al., 2014), tumor de células germinales (Kleiman et al., 2004), cáncer de colon (Tavakolian et al., 2020), leucemia linfocítica crónica y leucemia mieloide aguda (Depil et al., 2002). Además, la sobreexpresión o desregulación del gen *Pol* se ha visto asociada con cáncer de colon (Dolci et al., 2020) y leucemia (Bergallo et al., 2017). De igual manera, la sobreexpresión, desregulación, o sub expresión del gen *Env* ha reportado asociación con cáncer de próstata (Wallace et al., 2014), tumor de células germinales (Kleiman et al., 2004), melanomas (Büscher et al., 2006), cáncer de ovario (Wang-Johanning et al., 2007), cáncer gástrico, cáncer de colon (Tavakolian et al., 2020), leucemia linfoblástica aguda y leucemia mieloide aguda (Januszkiewicz-Lewandowska et al., 2013). Finalmente, la sobreexpresión, desregulación o sub expresión de las proteínas Rec y Np9 también se ha visto asociada con melanomas (Büscher et al., 2006), cáncer de ovario (Wang-Johanning et al., 2007), cáncer gástrico, cáncer de colon (Tavakolian et al., 2020), cáncer de mama (Tavakolian et al., 2019) y leucemia promielocítica (Denne et al., 2007).

El término leucemia puede asignarse a un grupo de enfermedades causadas por la alteración de los procesos reguladores celulares normales, que llevan a la proliferación descontrolada de células madre hematopoyéticas (glóbulos rojos, blancos y plaquetas) a la médula ósea (Davis et al., 2014). Esta enfermedad afecta a niños y adultos (adultos en mayor

proporción) y generalmente causa síntomas comunes como fiebre, sangrado o moretones, dolor muscular y óseo, fatiga y pérdida de peso, no obstante, algunos pacientes pueden ser asintomáticos (Davis et al., 2014). La leucemia es un problema a nivel mundial debido a que representa el 3,4% de los casos nuevos de cáncer, con un estimado de más de 60.000 pacientes nuevos en 2020, además, también representa el 3,8% de las muertes causadas por cáncer con un estimado de 23.000 muertes en 2020, y tiene una tasa de supervivencia relativa a 5 años del 63,7% (estadísticas tomadas en Estados Unidos) (Howlader et al., 2020).

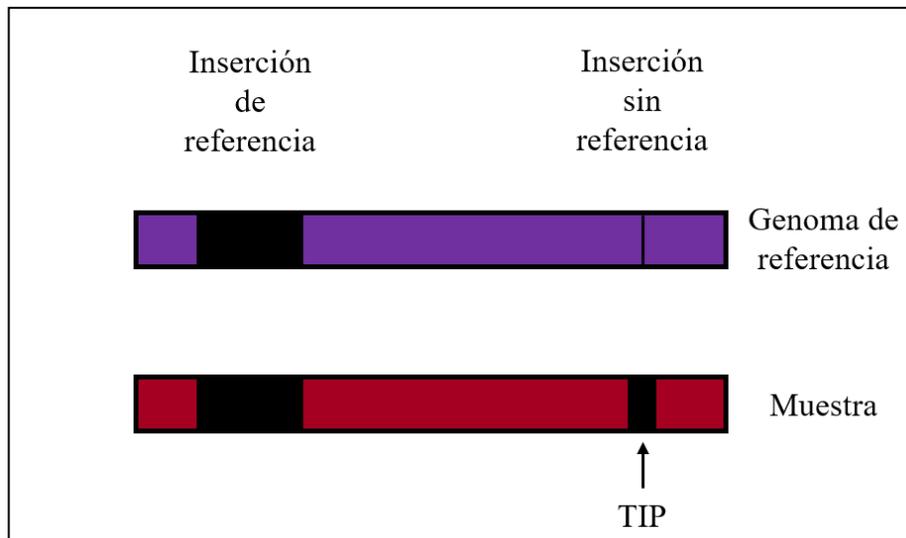
La leucemia puede clasificarse de dos maneras: dependiendo de la velocidad de desarrollo y progreso de la enfermedad, siendo aguda si los blastos (glóbulos blancos inmaduros) se reproducen rápidamente y crónica si los blastos se reproducen de manera más lenta; y dependiendo del tipo de glóbulo blanco afectado (células mieloides o células linfoides) (Howlader et al., 2020). Por lo tanto, los cuatro tipos de leucemia más comunes son: leucemia linfóide aguda (LLA), leucemia linfóide crónica (LLC), leucemia mielóide crónica (LMC) y leucemia mielóide aguda (LMA) (Davis et al., 2014)

La leucemia mielóide aguda (LMA) es definida como un grupo de desórdenes de las células madre hematopoyéticas clonales, que causan fallo en la diferenciación y proliferación anormal de éstas a la médula ósea, lo que lleva a una acumulación de células no funcionales llamadas mieloblastos (Stone et al., 2004). Además, la LMA presenta síntomas como fiebre, fatiga, pérdida de peso, sangrado y moretones y signos como hepatoesplenomegalia que es el aumento del tamaño del hígado y bazo y de forma menos común linfadenopatía que es el agrandamiento de los ganglios linfáticos (Davis et al., 2014). Este tipo de leucemia es el más común en adultos, comprendiendo el 1,1% de todos los casos nuevos de cáncer con casi 20.000 casos en 2020 y el 1,8% de las muertes causadas por cáncer con más de 11.000 muertes el año

pasado, con una tasa de supervivencia relativa a cinco años del 29,5% (Estadísticas tomadas en Estados Unidos) (Howlader et al., 2020; Leyto-Cruz, 2018).

Por otro lado, la LMA representa el 80% de las leucemias agudas en adultos mayores de 45 años (Davis et al., 2014), y en el 97% de los casos se han identificado alteraciones cromosómicas que generan trastornos en el proceso normal de maduración, proliferación y diferenciación celular de los precursores mieloides (Leyto-Cruz, 2018). Por lo tanto, para facilitar su diagnóstico y tratamiento, la Organización Mundial de la Salud (OMS), divide la leucemia mieloide aguda en varios grupos: LMA con anomalías genéticas (cambios genéticos o cromosómicos) que incluye translocaciones, inversiones, deleciones y fusiones de genes, LMA con cambios relacionados con mielodisplasia, LMA relacionada con administración previa de quimioterapia o radiación, LMA no especificada de otras maneras, sarcoma mieloide y proliferaciones mieloides relacionadas con el síndrome de Down (Leyto-Cruz, 2018; NIH National Cancer Institute, 2020).

Generalmente, la LMA es tratada dependiendo de si el paciente es apto para recibir la quimioterapia de inducción intensiva, la cual consiste en una infusión continua de citarabina con una antraciclina (generalmente daunorubicina). A continuación, si se llega al estado de remisión completa, se requiere una terapia pos-remisión, la cual consiste en quimioterapia convencional y un trasplante de células hematopoyéticas. Sin embargo, actualmente existen nuevas terapias enfocadas en diferentes procesos celulares como señalización a través de tirosinas quinasas u otras vías, regulación epigenética del ADN y la cromatina, exportación de proteínas y antígenos expresados en las células hematopoyéticas leucémicas, entre otras (Döhner et al., 2015).

Figura 3*Polimorfismo de inserción de elemento transponible (TIP)*

Como se mencionó anteriormente, la sobre expresión, desregulación o sub expresión de proteínas de HERV-K se ha visto asociada con diferentes tipos de cáncer, incluyendo la leucemia mieloide aguda (Bergallo et al., 2017; Colombo et al., 2018; Depil et al., 2002; Januszkiewicz-Lewandowska et al., 2013). Sin embargo, no se ha tenido en cuenta los polimorfismos de inserción de elementos transponibles (TIPs, por sus siglas en inglés), las posibles alteraciones que causan en el genoma y su relación con esta enfermedad. Un TIP (Figura 3) es una inserción de un elemento transponible en el genoma de un individuo, en una posición que no se presenta en el genoma de referencia, los cuales se han visto asociados con diferentes fenómenos biológicos como procesos de adaptación en *Oryza sativa* (Carpentier et al., 2019) y *Drosophila melanogaster* (Barrón et al., 2014) y regulación de la expresión génica, incluyendo modificación de la cromatina, empalmes alternativos y traducción de proteínas (Drongitis et al., 2019).

Debido a la cantidad de TIPs que pueden presentarse en un mismo individuo y el gran tamaño de los genomas, la identificación de estos en el genoma completo de un organismo

podría tomar cantidades de tiempo muy elevadas si se hiciera de manera *in vitro* en un laboratorio, debido a la complejidad asociada y la correspondiente cantidad masiva de datos. No obstante, los avances tecnológicos y computacionales han permitido que el análisis *in silico* resulte más eficiente en términos de tiempo y costo económico en comparación con análisis *in vivo* o *in vitro* (Mangul et al., 2019).

El análisis *in silico* en el contexto de las ciencias biológicas puede definirse como bioinformática, un conjunto de técnicas, estrategias y sistemas computacionales diseñados para el análisis de datos masivos de naturaleza biológica. Según Polanski & Kimmel, (2007), la bioinformática es una disciplina que surgió con el propósito de poner orden al conjunto de datos masivo que es producido por las nuevas tecnologías de biología molecular, además de la necesidad de herramientas para ensamblar secuencias y la anotación de estas secuencias. El análisis bioinformático ofrece múltiples aplicaciones, por ejemplo, la creación de bases de datos de ADN y proteínas, herramientas de comparación de secuencias, navegadores de genomas, predicción de transcritos de ARN, métodos de predicción de genes, anotación de genes, análisis de motivos regulatorios, predicción de estructura de proteínas, entre otros (D Edwards et al., 2009).

La unión de la bioinformática y la medicina facilita la creación de nuevas estrategias de terapia génica, diagnóstico temprano y predicción de múltiples enfermedades, mediante el análisis de genes, SNPs (*Single Nucleotide Polymorphism*), o incluso genomas completos (Chande et al., 2018; Y. Liu et al., 2019; Rong et al., 2018; Shao et al., 2019). En relación al cáncer, el uso de técnicas de bioinformática ha llevado a encontrar biomarcadores para terapia génica, por ejemplo, el uso de micro ARNs para el análisis de la patogénesis del carcinoma hepatocelular (Ding et al., 2017); la identificación de sellos distintivos del cáncer como

señalización sostenida de proliferación, resistencia a la muerte celular, angiogénesis inducida, entre otros (Lu et al., 2018); y marcadores para el diagnóstico y terapia del carcinoma nasofaríngeo (Xue et al., 2020), entre muchos otros casos de éxito.

Finalmente, con respecto a la leucemia mieloide aguda, herramientas de bioinformática han permitido analizar el perfil de expresión de ARNs largos no codificantes (lncRNAs, por sus siglas en inglés) y su relación con el sistema inmune y la diferenciación de las células hematopoyéticas en pacientes con LMA (Feng et al., 2018); caracterizar ARNs circulares como nuevos biomarcadores para el diagnóstico y tratamiento de la LMA (Wei Li et al., 2017); y la identificación de vías metabólicas, genes y micro ARNs claves para el diagnóstico y tratamiento de la enfermedad (S. Chen et al., 2020; Huang et al., 2017).

Otra aplicación de la bioinformática es la identificación de TIPs, para la cual se han utilizado diferentes softwares con el propósito de reducir la cantidad de tiempo utilizada en esta tarea. Entre los softwares más conocidos se encuentran TranspoSeq (Helman et al., 2014), Jitterbug (Hénaff et al., 2015) y Metasv (Mohiyuddin et al., 2015), los cuales en su mayoría se basan en el mapeo de extremos emparejados (PEM, por sus siglas en inglés) y lecturas divididas, en donde los métodos más comunes consisten en mapear todas las lecturas emparejadas en un genoma de referencia y las lecturas discordantes corresponden a variantes estructurales, las cuales son declaradas como TIP sí coinciden con un elemento transponible conocido (Carpentier et al., 2019; Ewing, 2015; Orozco-Arias et al., 2020).

No obstante, debido a que el paso del mapeo de las lecturas pareadas en el genoma de referencia es un paso muy demandante computacionalmente, Carpentier et al., (2019) desarrollaron el software TRACKPOSON, el cual emplea una metodología diferente que consiste en mapear todas las lecturas pareadas del individuo en cada familia de elementos

transponibles usando una secuencia consenso y a continuación, las lecturas discordantes son mapeadas en el genoma de referencia dividido en ventanas de 10 kb. Gracias a esto, TRACKPOSON permite identificar elementos transponibles rápidamente en conjuntos de datos grandes tales como el genoma del arroz (400 Mb) (Carpentier et al., 2019).

Sin embargo, este software no fue diseñado para ejecutarse en múltiples procesadores en paralelo y, por consiguiente, no es óptimo para analizar eficientemente genomas de mayor tamaño como el genoma humano, el cual está compuesto de 3.3 Gb. Consecuentemente, Orozco-arias et al., (2020), presentó el software TIP_finder, que utiliza la misma metodología del software TRACKPOSON e implementa, a su vez, técnicas de computación de alto rendimiento y programación en paralelo, esto con el propósito de permitir su ejecución al mismo tiempo en múltiples núcleos de CPU distribuidos en nodos (computadores), lo cual es conocido en este contexto como un supercomputador o clúster de alto rendimiento, permitiendo así, el análisis eficiente de genomas de gran tamaño. En este sentido, TIP_finder fue utilizado dentro de este proyecto para analizar la relación entre TIPS de HERV-K y leucemia mieloide aguda.

2.1.1 Método

2.1.1.1 Descripción del conjunto de datos.

Este proyecto de grado utilizó el conjunto de datos de los estudios L'Abbate et al., (2018) y J. Chen et al., (2019) y el bioproyecto Decitabine del SRA con código de acceso PRJNA242906, los cuales consisten en 16 pacientes con leucemia mieloide aguda y cariotipo normal, 17 pacientes con leucemia mieloide aguda y alteraciones cromosómicas (Apéndice A) y 33 pacientes sanos (controles) (Apéndice B); el criterio para la selección de los pacientes control fue la no existencia de enfermedades diagnosticadas. El conjunto de datos comprende 33 pacientes femeninos y 33 masculinos en un rango de edad de 20 a 90 años. Estos datos fueron

encontrados al realizar una búsqueda en la base de datos SRA (Leinonen et al., 2011) del NCBI con las siguientes cadenas de búsqueda: “((((((((((Acute Myeloid Leukemia) AND Homo sapiens[Organism]) AND PAIRED[Layout])) AND GENOMIC[Source])) AND WGS[Strategy])) NOT exome)) AND public [Access]” para los pacientes de LMA y “((((((((((Homo sapiens[Organism]) AND PAIRED[Layout])) AND GENOMIC[Source])) AND WGS[Strategy])) NOT exome)) AND public [Access]” para los controles. Esto con el objetivo de filtrar por condición (LMA o control), *Homo sapiens* como organismo, secuenciación de genoma completo pareada con Illumina (Voelkerding et al., 2009) y obtener datos de acceso público.

A continuación, se descargaron los datos del SRA a través del programa fastq-dump (Sequence Read Archive Submissions Staff, 2011) tomando máximo 30 millones de lecturas, con el fin de reducir la cantidad de datos a aproximadamente 1 TB (debido a limitaciones de almacenamiento del supercomputador utilizado) y, además, eliminar sesgos relacionados a la profundidad debido a que los datos con cariotipo normal contienen menos de 30 millones de lecturas. Posteriormente, se construyó la librería de HERV-K (Apéndice C) usando las bases de datos del Centro Nacional para la Información Biotecnológica (NCBI, por sus siglas en inglés) (NCBI Resource Coordinators, 2016), el Archivo Europeo de Nucleótidos (ENA, por sus siglas en inglés) (Leinonen et al., 2011) y el Banco de Datos de ADN de Japón (DDBJ, por sus siglas en inglés) (Mashima et al., 2017) con las siguientes cadena de búsqueda: “(“Human endogenous retrovirus K”[Organism] OR HERV-K[All Fields]) AND “Homo sapiens”[porgn] AND (biomol_genomic[PROP] AND is_nucore[filter] AND (“5000”[SLEN] : “15000”[SLEN]))”, “HERV-K AND Homo sapiens” filtrando por tamaño de secuencia (5000-15000) y “SequenceLength:[5000 TO 15000]HERV-K AND Homo sapiens”, respectivamente. El rango

de longitud de la secuencia fue seleccionado debido a que el tamaño promedio aproximado de un retrovirus endógeno humano es de 10 kb (Shin et al., 2013). Finalmente, se descargó el genoma humano de referencia versión 38 del NCBI y su anotación correspondiente.

2.1.1.2 Identificación de los TIPS de HERV-K

Los TIPS de HERV-K en el genoma de pacientes de LMA y controles, fueron identificados con el software TIP_finder (Orozco-Arias et al., 2020), un conjunto de componentes software interconectados (*pipeline*), desarrollado en el lenguaje de programación Python (versión 3.8.7), como se describe en la tabla 1. Los datos de entrada de TIP_finder son archivos en formato fastq de pacientes y controles, con máximo 30 millones de lecturas forward y reverse. Los análisis realizados utilizaron los recursos del supercomputador Itrop (<https://bioinfo.ird.fr/>) del Instituto Francés de Investigación para el Desarrollo (IRD, por sus siglas en francés).

Tabla 1

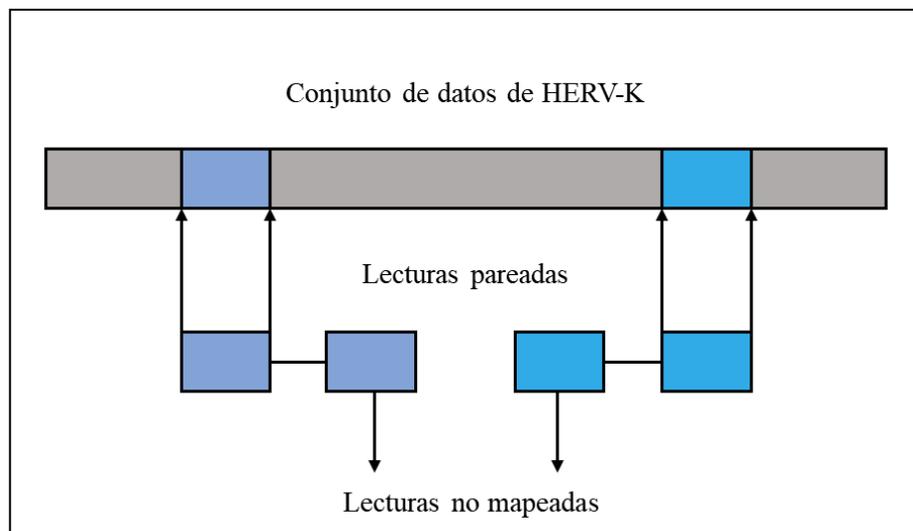
Pipeline de TIP_finder

Software	Función	Versión	Referencia
Blast	Alineamiento de secuencias de tipo local ya sea de ADN, ARN o proteínas.	2.5.0	(Altschul et al., 1990)
Bowtie2	Alineamiento de grandes conjuntos de secuencias cortas o lecturas de ADN. Puede usarse en varios procesadores simultáneamente para mayor velocidad.	2.4.1	(Langmead & Salzberg, 2012)
Samtools	Conjunto de utilidades para interactuar y post procesar alineamientos de lecturas de secuencias de ADN, tales como: indexación, llamador de variantes y visor de alineamientos.	1.10	(H. Li et al., 2009)
Bedtools	Comparación, manipulación y anotación de características genómicas y alineamientos de secuencias.	2.29.2	(Quinlan & Hall, 2010)
Mpi4py	Es un envoltorio en Python para el estándar MPI (Message Passing Interface) que facilita la ejecución de procesos en paralelo en múltiples núcleos de CPU distribuidos en los nodos de un supercomputador.	Última versión	(Dalcin et al., 2011)

El pipeline de TIP_finder consiste en tres pasos, el primero mapea las lecturas emparejadas contra la librería de HERV-K y mantiene las lecturas en donde solo una lectura es mapeada y la otra no, también llamadas lecturas discordantes (Figura 4), conservando las que no se mapearon. A continuación, las lecturas no mapeadas se alinean con el genoma de referencia para encontrar los puntos de inserción de cada TIP; el segundo, procesa los alineamientos para buscar las lecturas con solo un acierto y genera un archivo tabular (formato bed); finalmente, este archivo es usado para contar cuantas lecturas se alinearon con el genoma de referencia (dividido previamente en fragmentos de 10 kb) y genera un archivo delimitado tabular para cada condición (Apéndices D y E) los cuales son procesados por el script de Python (versión 3.8.7) TIP_finder_utils.py (Orozco-Arias et al., 2020)

Figura 4

Lecturas discordantes



Este script se encarga de procesar y analizar los datos generados de la siguiente manera: primero, agrupa los resultados en una matriz booleana de TIPs para cada archivo de secuencias de extremos emparejados (Paired-end sequences); a continuación, genera gráficos de frecuencia

del número de TIPS de cada muestra analizada; posteriormente, realiza una tabla de picos usando la frecuencia de TIPS presentes en cada cromosoma; y finalmente, lleva a cabo pruebas de asociación estadística para identificar los TIPS con el mayor rango de asociación a la leucemia mieloide aguda (Orozco-Arias et al., 2020).

2.1.1.3 Análisis estadístico

El análisis de correlación entre TIPS de HERV-K y leucemia mieloide aguda se basó en el enfoque descrito por Orozco-arias et al., (2020), el cual consiste en analizar los siguientes indicadores estadísticos: primero, se establecen dos variables categóricas para determinar si el paciente i tiene LMA o no ($i=1\dots n$) y para determinar si el paciente i tiene el TIPS j o no ($j=1\dots m$), n cantidad de pacientes y m cantidad de TIPS. A continuación, usando las variables mencionadas anteriormente, se crea una tabla de contingencia para cada TIPS identificado, a las cuales se les realiza una prueba de hipótesis de independencia no paramétrica basada en el test de distribución chi cuadrado (χ^2), con un grado de libertad y un $\alpha = 0.05$, con el objetivo de determinar si el TIPS j tiene una influencia estadísticamente significativa en la leucemia mieloide aguda. Finalmente, una vez identificados los TIPS asociados a LMA, se evaluó el nivel de asociación en un rango de 0 a 1, calculando el coeficiente de contingencia de χ^2 (Cochran, 1954).

2.1.1.4 Identificación de posibles genes afectados por los TIPS

Los TIPS asociados fueron filtrados con un P-valor menor a 1×10^{-5} . A continuación, para visualizar las regiones de interés de cada cromosoma se utilizó la librería de R shinyCircos (Yu et al., 2018), la cual requiere tres archivos separados por comas (.csv) para su funcionamiento, los cuales son: un archivo con el tamaño de cada cromosoma del genoma humano y un archivo con la cantidad de TIPS encontrados en una ventana de 10 Mb (esto para facilitar la visualización

de los TIPS en cada cromosoma) para cada condición (LMA o control). El conteo de los TIPS encontrados para cada condición fue realizado con un script de Python (versión 3.8.7).

Posteriormente, se seleccionó el cromosoma 8 basándose en la abundante presencia de TIPS en pacientes de LMA y la ausencia de TIPS en controles en la misma región. Finalmente, se observó el cromosoma 8 en shinyCircos utilizando los siguientes archivos: un archivo con el tamaño del cromosoma 8 del genoma humano, un archivo con la cantidad de TIPS encontrados en una ventana de 10 kb para cada condición (LMA o control) y un archivo con los TIPS asociados que contiene la ubicación y el porcentaje de asociación.

Una vez se seleccionó la región de interés del cromosoma 8 (125.000.000-130.000.000), se procedió a observar la región en el navegador de genoma de Ensembl (Yates et al., 2020). A continuación, se observó si había presencia de un TIP dentro o cerca de una región de interés de un gen (codificante o no). Para esto se agregó un carril personalizado en el navegador de genoma utilizando un archivo separado por tabulaciones (.bed) con la posición de cada TIP. Finalmente, una vez seleccionados los genes de interés se procedió a realizar la búsqueda bibliográfica de relaciones del gen con la leucemia mieloide aguda.

2.1.2 Resultados

La leucemia mieloide aguda es un problema a nivel mundial debido a que es el tipo más común en adultos, representando el 1,1% de todos los casos nuevos de cáncer y una tercera parte de todas las leucemias, con una tasa de supervivencia relativa a cinco años del 29,5%. Por esta razón, es importante diseñar estrategias de terapia génica que sean más eficientes y menos dañinas que la quimioterapia. No obstante, el diseño de métodos de diagnóstico y predicción que ayuden a las personas a recibir tratamiento de manera más eficiente y cuando la enfermedad no está tan avanzada como para mostrar signos, es otra parte muy importante para contribuir a la

remisión y disminución de muertes causadas por esta enfermedad. Por consiguiente, los resultados de este proyecto muestran la asociación estadística de TIPS de HERV-K con leucemia mieloide aguda usando lecturas de secuenciación de genomas completos de pacientes y controles, como se muestra a continuación.

Tabla 2

TIPs de HERV-K asociados a leucemia mieloide aguda

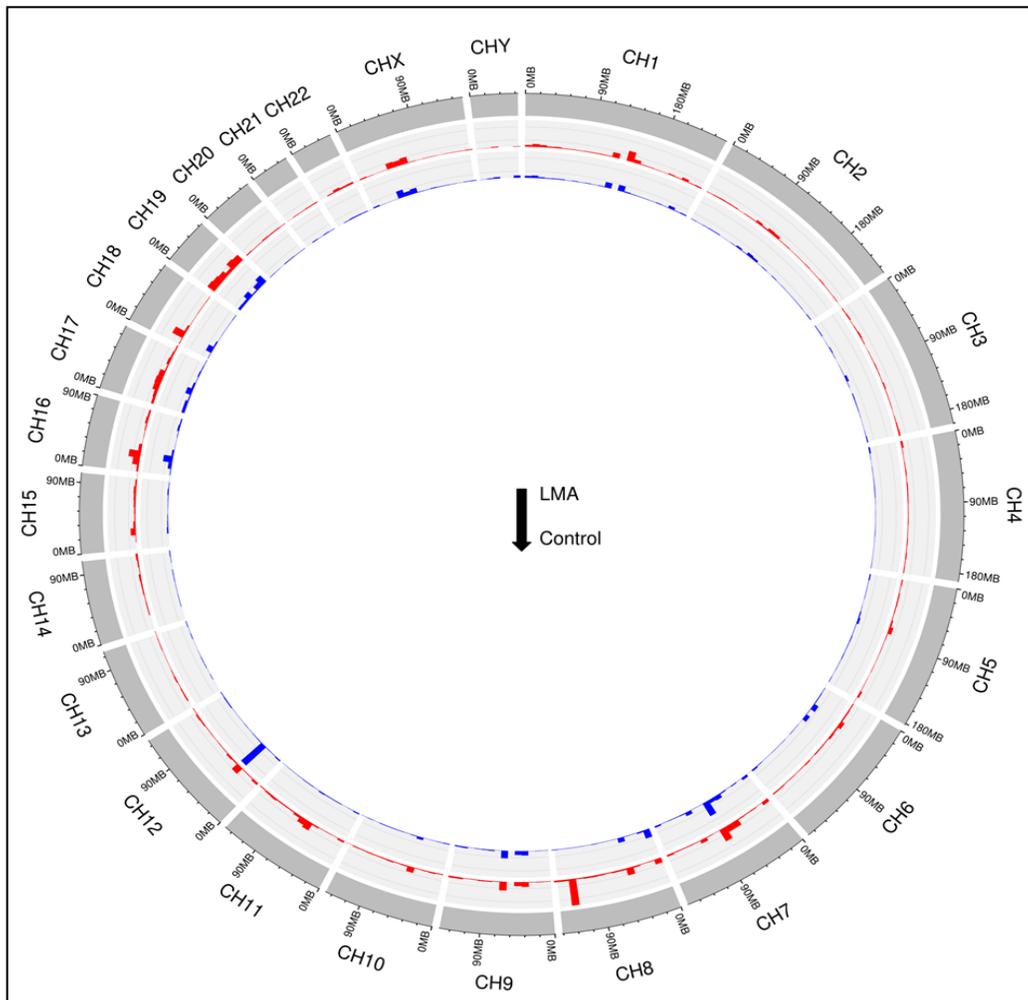
Posición	LMA	Control	P-valor	Asociación (%)
Ch16:14770000-14780000	26	0	$3,9661 \times 10^{-8}$	56,60
Ch8:125460000-125470000	25	0	$8,7513 \times 10^{-8}$	55,60
Ch8:128110000-128120000	25	0	$8,7513 \times 10^{-8}$	55,60
Ch16:16400000-16410000	25	0	$8,7513 \times 10^{-8}$	55,60
Ch8:127840000-127850000	25	0	$8,7513 \times 10^{-8}$	55,60
Ch8:126780000-126790000	24	0	$1,9039 \times 10^{-7}$	54,56
Ch8:129560000-129570000	24	0	$1,9039 \times 10^{-7}$	54,56
Ch8:127830000-127840000	26	1	$1,9967 \times 10^{-7}$	54,50
Ch8:125400000-125410000	26	1	$1,9967 \times 10^{-7}$	54,50
Ch8:125480000-125490000	23	0	$4,0866 \times 10^{-7}$	53,49

Se muestran 10 de 101 TIPs asociados con LMA y filtrados con un P-valor menor a 1×10^{-5}

La tabla 2 muestra los TIPs de HERV-K con mayor asociación estadística con la leucemia mieloide aguda. Las columnas LMA y control muestran la cantidad de individuos que presentan el TIP en esa posición, por ejemplo, 26 de 33 pacientes de LMA presentan el TIP y ningún control presenta un TIP en esa posición. Además, se muestra el P-valor y el porcentaje de asociación de cada TIP con la LMA (Ver apéndice F para resultados completos).

Figura 5

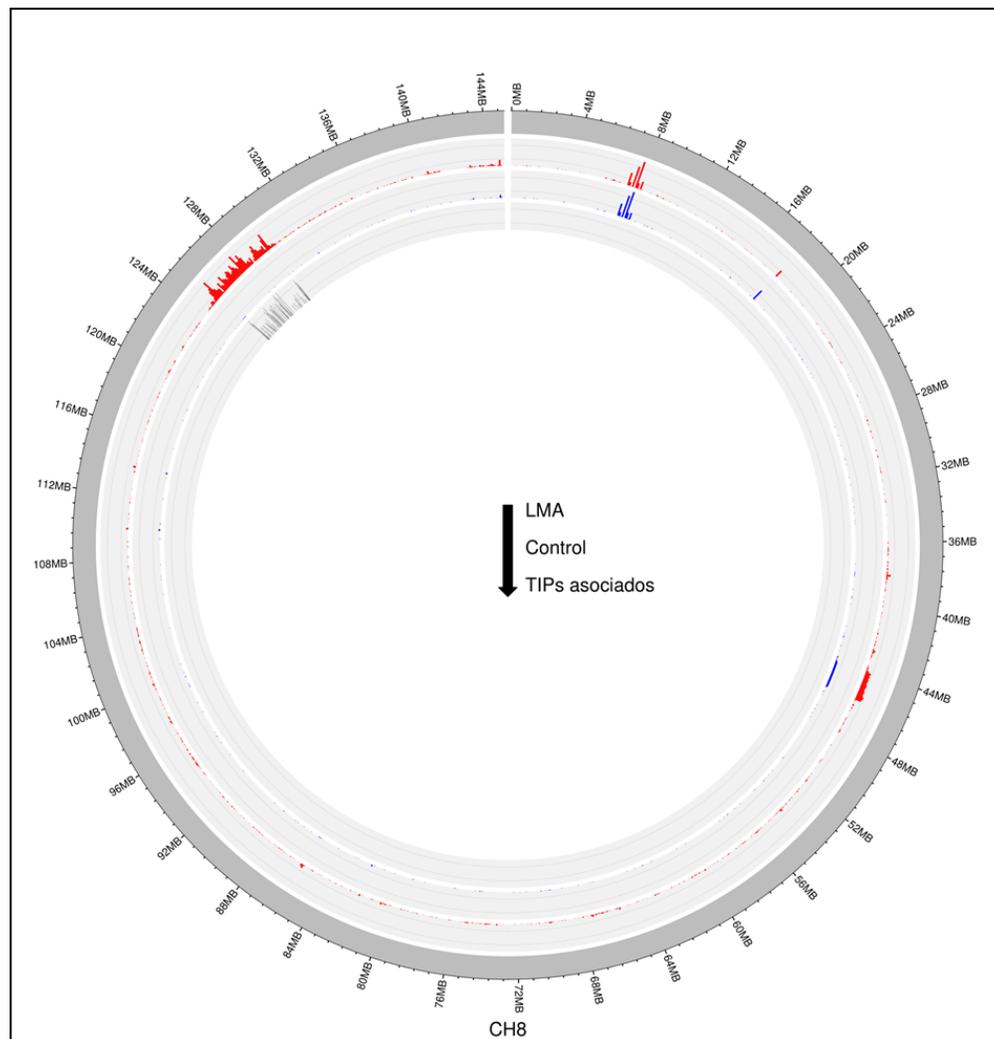
TIPs de HERV-K en el genoma completo de pacientes de LMA y controles



La figura 5 muestra el genoma completo de pacientes de LMA (carril externo) y controles (carril interno) divididos en cromosomas y cada cromosoma dividido en ventanas de 10 Mb. Además, se muestran los correspondientes TIPs encontrados en cada cromosoma. Se observan varios picos de TIPs en diferentes cromosomas, tanto en los pacientes (en rojo) como en controles (en azul); no obstante, el cromosoma 8 es el que presenta el nivel más elevado de TIPs en pacientes con LMA y no se observa el mismo resultado en los controles.

Figura 6

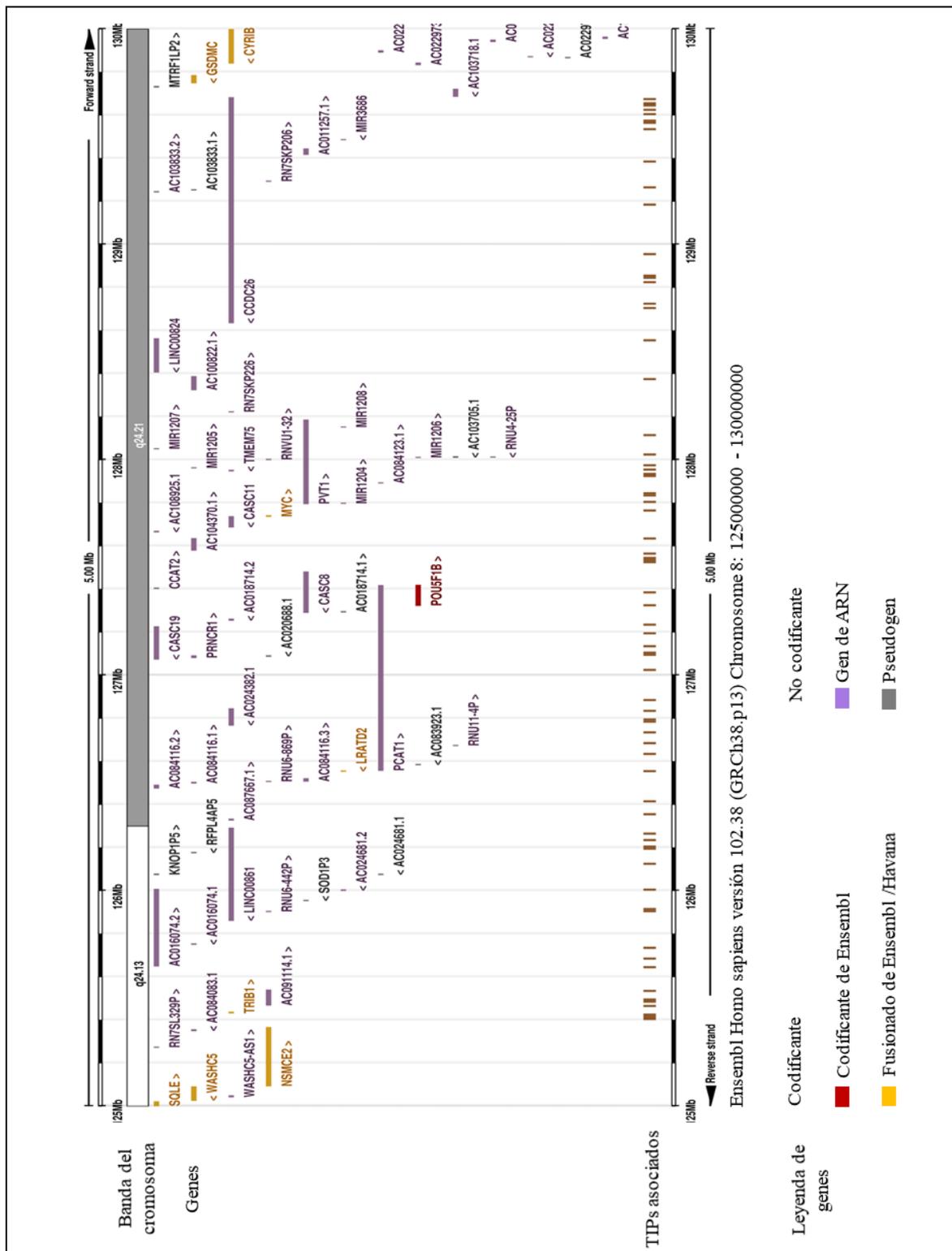
TIPs de HERV-K asociados a LMA en el cromosoma 8 de pacientes de LMA y controles



Por esta razón, se hace un acercamiento al cromosoma 8 dividido en ventanas de 10 kb, como se muestra en la figura 6, en donde se pueden observar los dos carriles mencionados anteriormente y un carril extra en donde se muestran los TIPs de HERV-K (en negro) asociados a LMA encontrados en esta región para pacientes y controles. Como se mencionó anteriormente, los TIPs fueron filtrados con un P-valor menor a 1×10^{-5} .

Figura 7

Navegador de genoma para la región de interés del cromosoma 8 (125.000.000-130.000.000)



Debido a la cantidad de TIPS encontrados en la región de la base 125.000.000 a la 130.000.000 del cromosoma 8, se procedió a observar esta región en el navegador de genoma de Ensembl, como se observa en la figura 7, en donde se pueden observar los genes codificantes, no codificantes y los TIPS de HERV-K asociados a LMA encontrados en esta región. Por consiguiente, con el objetivo de ofrecer un mayor nivel de detalle a los genes presentados en la figura 6, se presenta la siguiente tabla:

Tabla 3

Características de los genes encontrados en la región de interés

Nombre	Tipo	Función	Posición	Presencia de TIP	Posición de TIP	Link	Referencia
<i>SQLE</i>	Codificante	Cataliza el primer paso de oxigenación en la biosíntesis de esteroides y se cree que es una de las enzimas que limitan la velocidad en esta vía. Asociado con cáncer de pulmón, de mama, carcinoma hepatocelular, cáncer de próstata y leucemia mieloide aguda	124,998,497 - 125,022,283 Forward	No	NA	https://www.ncbi.nlm.nih.gov/genome/6713	(Brown et al., 2016; Ge et al., 2019; Ma et al., 2021)
<i>WASHC5</i>	Codificante	Este gen codifica una proteína de 134 kDa llamada strumpellin que se predice que tiene múltiples dominios transmembrana y un dominio que contiene repeticiones de espectrina. Asociado con leucemia mieloide aguda.	125,024,260 - 125,091,819 Reverse	No	NA	https://www.ncbi.nlm.nih.gov/genome/9897	(Guo et al., 2020)
<i>NSMCE2</i>	Codificante	Este gen codifica un miembro de una familia de ligasas modificadoras pequeñas relacionadas con ubiquitina (SUMO) E3 que median la unión de una proteína SUMO a proteínas involucradas en el transporte nuclear, la transcripción, la segregación cromosómica y la reparación del ADN. Asociado con leucemia mieloide aguda.	125,091,679 - 125,367,120 Forward	No	NA	https://www.ncbi.nlm.nih.gov/genome/286053	(Chinen et al., 2014; L'Abbate et al., 2018)
<i>TRIB1</i>	Codificante	Actividad transferasa, transferencia de grupos que contienen fósforo y unión al factor de transcripción. Proteína adaptadora involucrada en la degradación de proteínas al interactuar con la ubiquitina ligasa COP1. Asociado a cáncer de mama, de próstata y leucemia mieloide aguda.	125,430,358 - 125,438,403 Forward	Si	Flanco del promotor, promotor.	https://www.ncbi.nlm.nih.gov/genome/10221	(Keeshan et al., 2016; Liang et al., 2013; Nakamura, 2015; Yokoyama et al., 2010)

LRATD2	Codifica nte	Proteína asociada a la membrana del cáncer de mama. Asociado con cáncer gástrico, próstata, y de ovario, otros cánceres.	126,552,443 - 126,558,478 Reverse	Si	Flanco del promotor, promotor.	https://www.ncbi.nlm.nih.gov/genbank/157638	(Cheng et al., 2016; Wong et al., 2017; Y. Zhang et al., 2019)
POU5F1 B	Codifica nte	Puede codificar una proteína que tiene casi la misma longitud y es muy similar al factor de transcripción POU clase 5 homeobox 1, se ha demostrado que es un activador transcripcional débil y puede desempeñar un papel en la carcinogénesis y el desarrollo ocular. Asociado con múltiples tipos de cáncer incluyendo leucemia mieloide aguda.	127,322,183 - 127,420,066 Forward	Si (2)	Flanco del promotor.	https://www.ncbi.nlm.nih.gov/genbank/5462	(Hayashi et al., 2015; Kastler et al., 2010; Pan et al., 2018; Yi et al., 2019; Zhu et al., 2020)
MYC	Codifica nte	Este gen es un protooncogén y codifica una fosfoproteína nuclear que juega un papel en la progresión del ciclo celular, la apoptosis y la transformación celular. Asociado con múltiples tipos de cáncer incluyendo leucemia mieloide aguda.	127,735,434 - 127,742,951 Forward	No	TIP a 17kb aproximadamente.	https://www.ncbi.nlm.nih.gov/genbank/4609	(Mughal et al., 2017; Ohanian et al., 2019; Yun et al., 2019)
MTRF1L P2	Pseudo gen	-	129,728,744 - 129,730,445 Forward	No	NA	https://www.ncbi.nlm.nih.gov/genbank/?term=MTRF1LP2	-
GSDMC	Codifica nte	Esta forma constituye el precursor de la proteína formadora de poros: tras la escisión, el resto N-terminal liberado (Gasdermin-C, N-terminal) se une a las membranas y forma poros, lo que desencadena la muerte celular. Asociado con cáncer colorrectal y leucemia mieloide aguda.	129,748,196 - 129,786,624 Reverse	No	NA	https://www.ncbi.nlm.nih.gov/genbank/56169	-
CYRIB	Codifica nte	Regula la quimiotaxis, la migración celular y la polarización epitelial controlando la polaridad, plasticidad, duración y extensión de las protuberancias. Asociado con cáncer de vesícula biliar y leucemia mieloide aguda.	129,839,593 - 130,017,129 Reverse	No	NA	https://www.ncbi.nlm.nih.gov/genbank/51571	(L'Abbate et al., 2018)
AC09111 4.1	No codifica nte LncRN A	Involucrado en rasgos como el ancho de la distribución de los glóbulos rojos. Recuento de basófilos, recuento de glóbulos blancos. Porcentaje de neutrófilos de leucocitos. Entre otros	125,466,939 - 125,541,373 Forward	Si (3)	Flanco del promotor, potenciador, CTCF.	https://www.ebi.ac.uk/genomes/AC091114.1	-

AC01607 4.2	No codifica nte LncRN A	Porcentaje de eosinófilos y neutrófilos y monocitos en glóbulos blancos. Ancho de la distribución de los glóbulos rojos.	125,648,327 - 126,008,930 Forward	Si (4)	Flanco del promotor, potenciador, CTCF.	https://www.ncbi.nlm.nih.gov/genomes/AC016074.2	-
PCATI	No codifica nte LncRN A	Este gen produce un ARN largo no codificante que promueve la proliferación celular. Asociado con varios tipos de cáncer incluyendo leucemia mieloide aguda.	126,556,323 - 127,419,050 Forward	Si (14)	Todas las regiones.	https://www.ncbi.nlm.nih.gov/genome/100750225	(Ghafouri-Fard et al., 2020; X.-L. Liu et al., 2019; P. Zhang et al., 2019)
AC02438 2.1	No codifica nte LncRN A	Asociado con cáncer de próstata	126,766,196 - 126,846,981 Reverse	Si (2)	CTCF, cromatina abierta, Flanco del promotor, potenciador.	https://www.ncbi.nlm.nih.gov/genomes/AC024382.1	-
CASC19	No codifica nte LncRN A	Asociado con cáncer gástrico, de pulmón, próstata y colorrectal.	127,072,694 - 127,227,541 Reverse	Si (4)	Flanco del promotor, potenciador, cromatina abierta	https://www.ncbi.nlm.nih.gov/genome/103021165	(Qu et al., 2019; Teerlink et al., 2016; W.-J. Wang et al., 2019)
PRNCRI	No codifica nte LncRN A	Asociado con cáncer de pulmón, colorrectal, gástrico, de mama y próstata	127,079,874 - 127,092,600 Forward	Si	Potenciador, cromatina abierta.	https://www.ncbi.nlm.nih.gov/genome/101867536	(AlMutairi et al., 2019; Hong et al., 2019; N. Li et al., 2021)
CASC8	No codifica nte LncRN A	Asociado con leucemia linfoblástica aguda, cáncer de vejiga, pulmón, gástrico y de próstata	127,289,817 - 127,482,139 Reverse	Si (2)	Flanco del promotor.	https://www.ncbi.nlm.nih.gov/genome/727677	(Hashemi et al., 2016; L. Hu et al., 2016; R. Hu et al., 2017)
AC10437 0.1	No codifica nte LncRN A	Asociado con cáncer de ovario epitelial, próstata y colorrectal	127,578,473 - 127,636,867 Forward	Si	Cromatina abierta.	https://www.ncbi.nlm.nih.gov/genomes/AC104370.1	-
PVTI	No codifica nte LncRN A	Es un candidato potencial a oncogén, debido a que la sobreexpresión de este se ha visto asociada con múltiples cánceres incluyendo leucemia mieloide aguda	127,794,526 - 128,187,101 Forward	Si (7)	Flanco del promotor, CTCF, potenciador, promotor, cromatina abierta.	https://www.ncbi.nlm.nih.gov/genome/5820	(Chinen et al., 2014; Dieter et al., 2020; El-Khazragy et al., 2019; Izadifard et al., 2018)

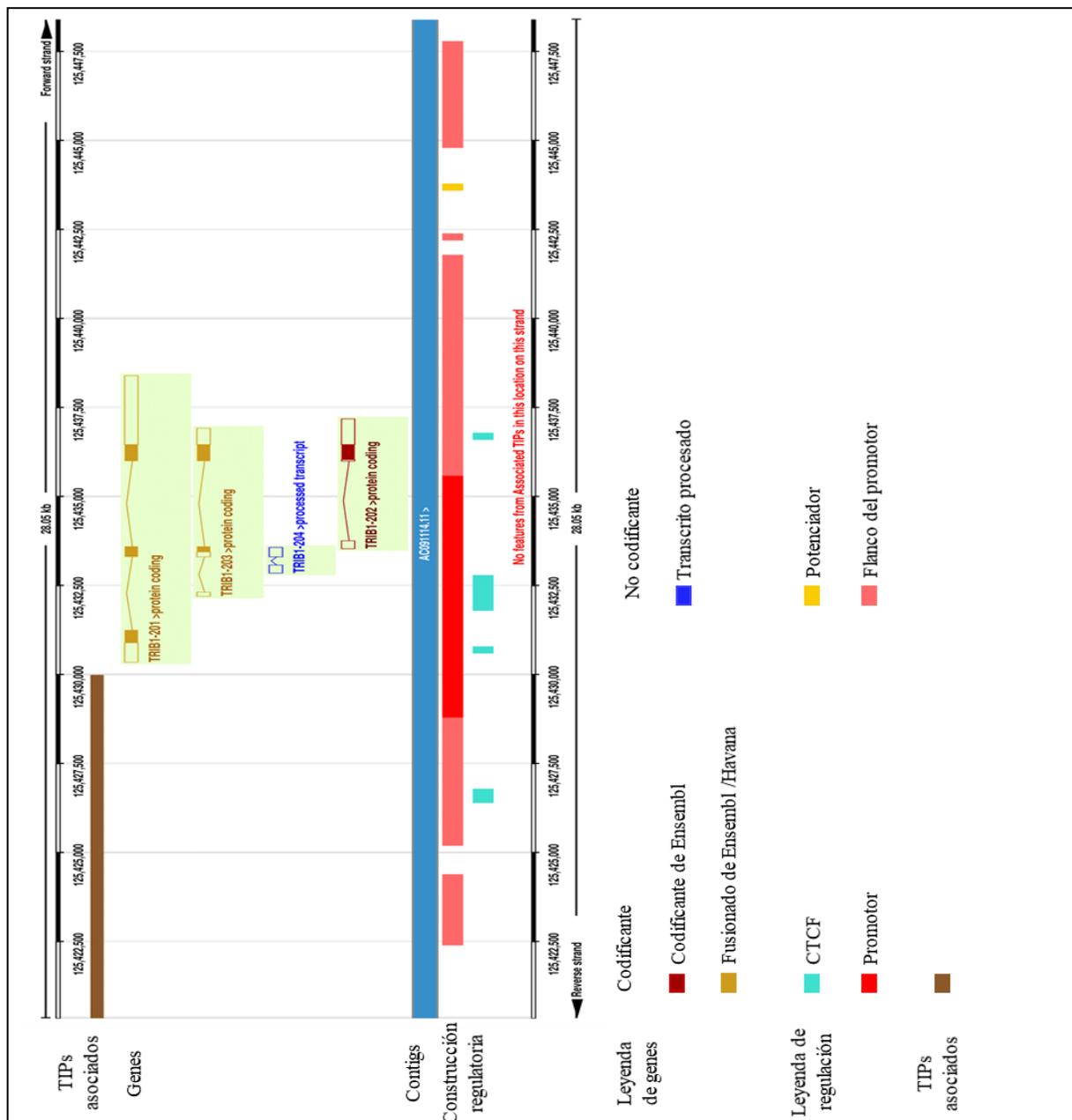
AC10082.1	No codificante LncRN A	Asociado con linfoma difuso de células B grandes	128,323,131 - 128,388,636 Forward	Si	Flanco del promotor, cromatina abierta, CTCF.	https://www.ebi.ac.uk/genwas/genes/AC10082.1	-
LINC00824	No codificante LncRN A	Asociado con cáncer de ovario, páncreas, pulmón, endometrio,	128,405,269 - 128,564,679 Reverse	Si	CTCF, cromatina abierta.	https://www.ebi.ac.uk/genwas/genes/LINC00824	(Burghaus et al., 2017; He et al., 2020)
CCDC26	No codificante LncRN A	Asociado con cáncer de páncreas, glioma y leucemia mieloide aguda	128,634,199 - 129,683,770 Reverse	Si (14)	Todas las regiones excepto promotor.	https://www.ncbi.nlm.nih.gov/genbank/137196	(C. Chen et al., 2019; Dieter et al., 2020; Hirano et al., 2015; Izadifard et al., 2018)

CTCF refiere a los sitios de unión CTCF los cuales son un factor de transcripción.

Como puede observarse, la tabla 3 incluye información como el tipo de gen, función, posición, presencia y posición del TIP, link del gen y referencias, sobre todos los genes codificantes (10) encontrados en la región de interés y 12 genes no codificantes, los cuales fueron seleccionados debido a la presencia de uno o más TIPs en alguna de sus regiones de interés. De los 22 genes encontrados en la tabla 3 cabe destacar cuatro genes codificantes: *TRIB1*, *LRATD2*, *POU5F1B* y *MYC*, los cuales fueron seleccionados debido a la presencia de uno o más TIPs en sus regiones de interés, con la excepción del gen *MYC*, el cual fue seleccionado debido a su asociación con múltiples tipos de cáncer incluyendo leucemia mieloide aguda; y 3 genes no codificantes: *PCATI*, *PVT1* y *CCDC26*, debido a su relación con la leucemia mieloide aguda en la literatura, los cuales se muestran a continuación:

Figura 8

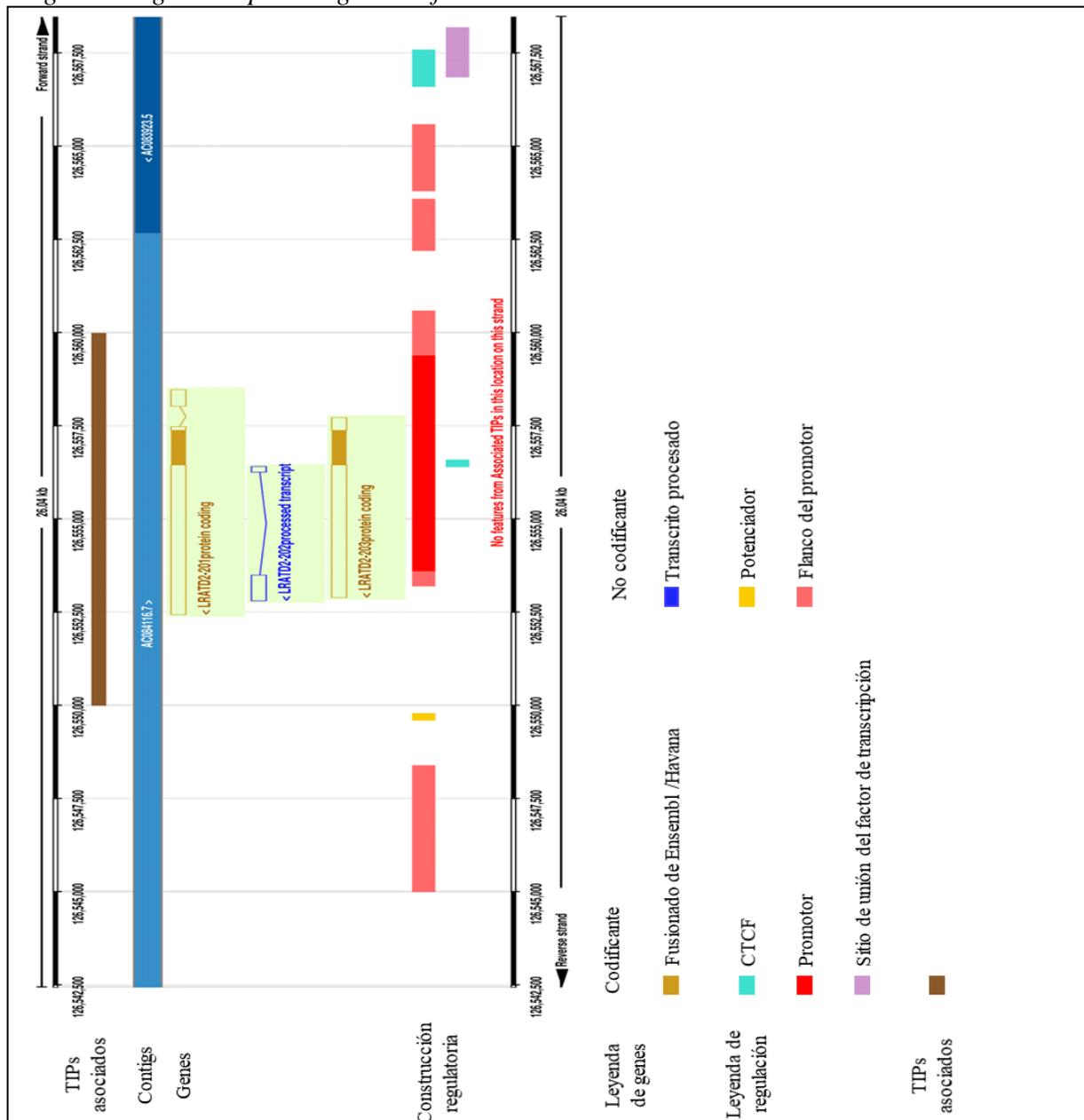
Navegador de genoma para el gen codificante *TRIB1*



En primer lugar, la figura 8 muestra el navegador de genoma para el gen codificante *TRIB1* con sus regiones de interés y su respectiva leyenda. Además, muestra la posición del TIP de HERV-K asociado a LMA. Se puede observar que el TIP está ubicado en las regiones del flanco del promotor y el promotor del gen.

Figura 9

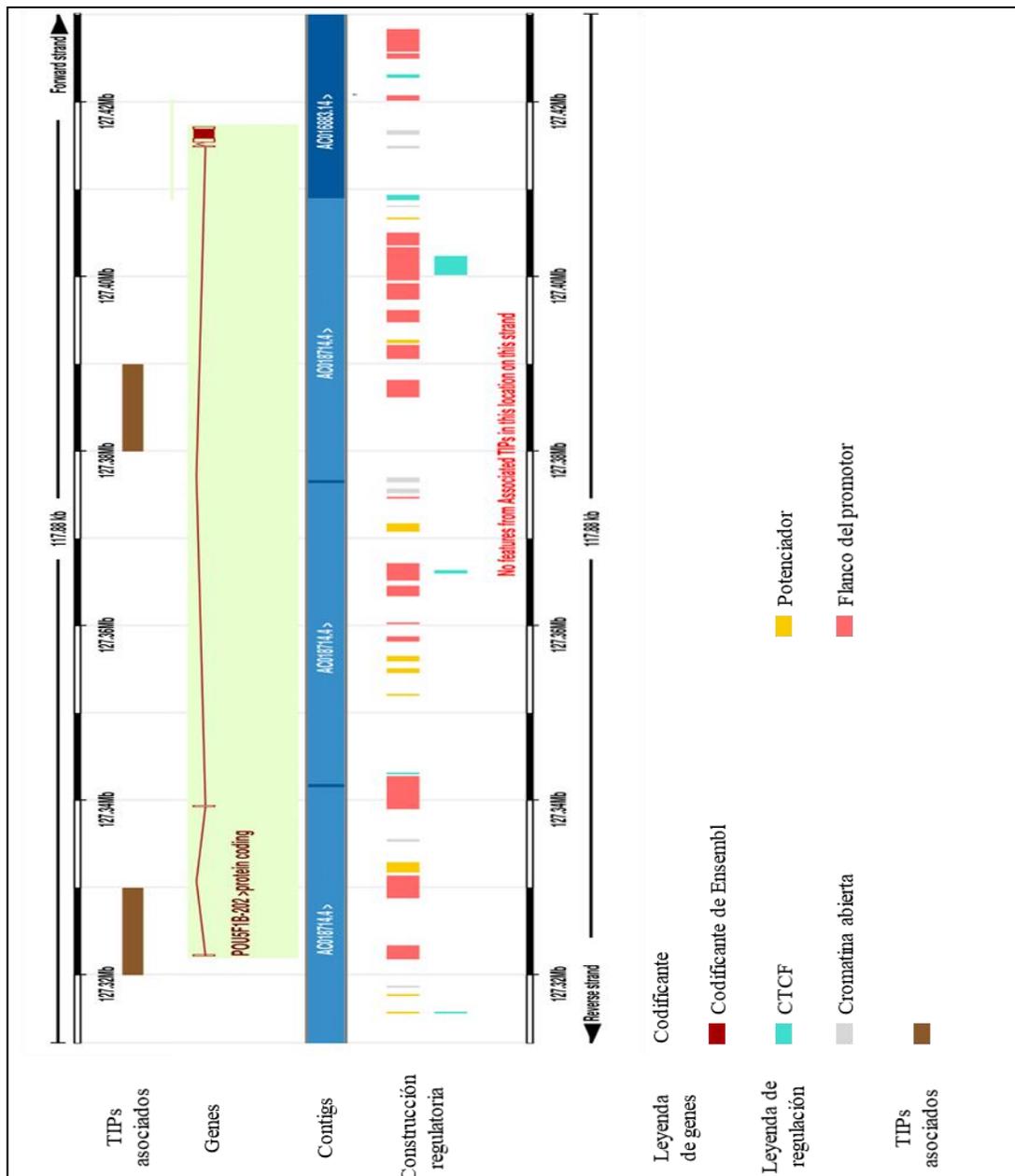
Navegador de genoma para el gen codificante *LRATD2*



A continuación, se presenta el navegador de genoma para el gen codificante *LRATD2* con sus regiones de interés y su respectiva leyenda como se observa en la figura 9. Además, muestra la posición del TIP de HERV-K asociado a LMA. Se puede observar que este TIP superpone al gen y, además, está ubicado sobre las regiones del flanco del promotor, promotor y CTCF.

Figura 10

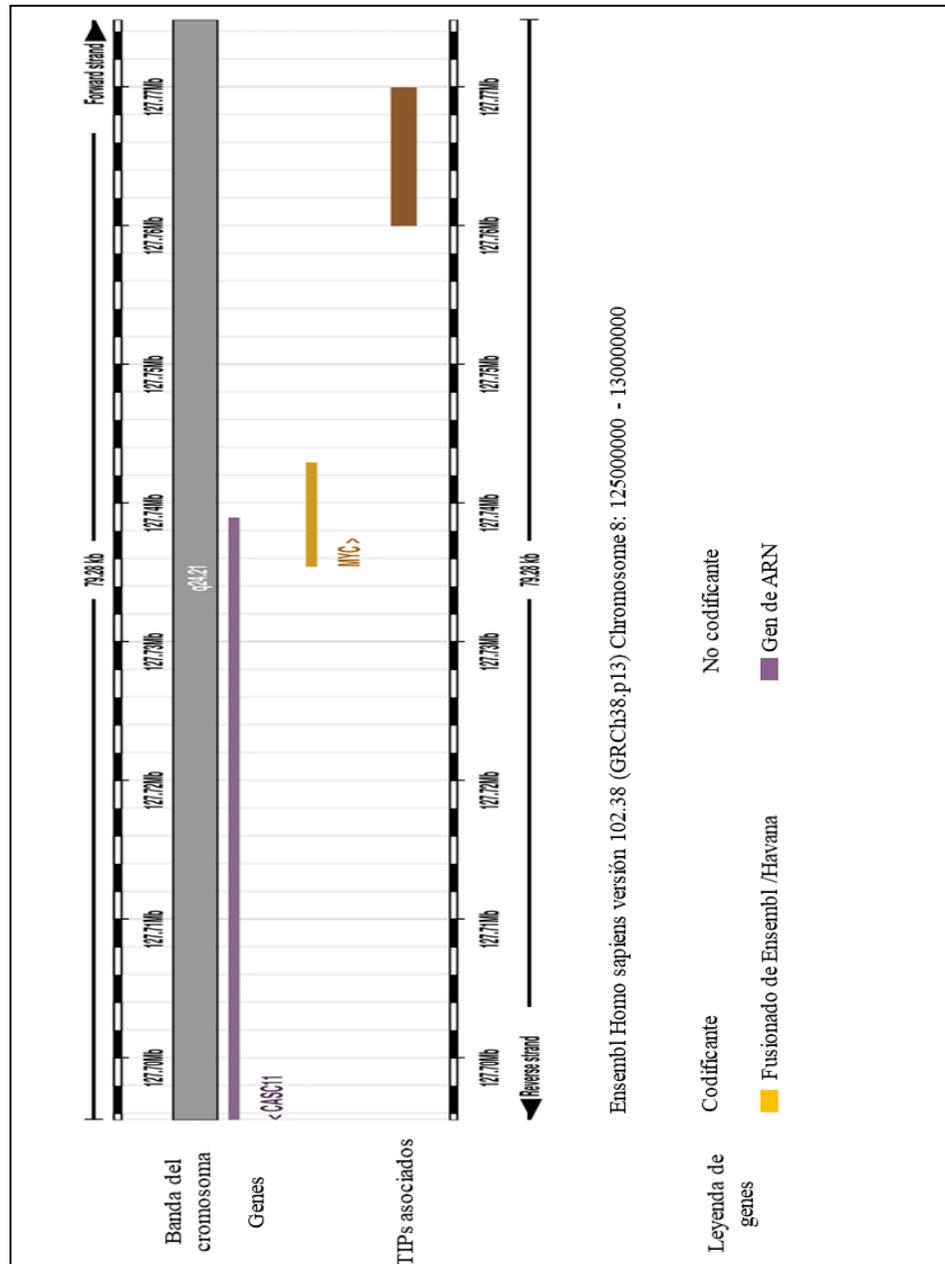
Navegador de genoma para el gen codificante POU5F1B



De igual manera, la figura 10 muestra el navegador de genoma para el gen codificante *POU5F1B* con sus regiones de interés y su respectiva leyenda. Además, muestra las posiciones de los TIPS de HERV-K asociados a LMA. Se puede observar que los TIPS superponen al gen y además están ubicados sobre la región del flanco del promotor.

Figura 11

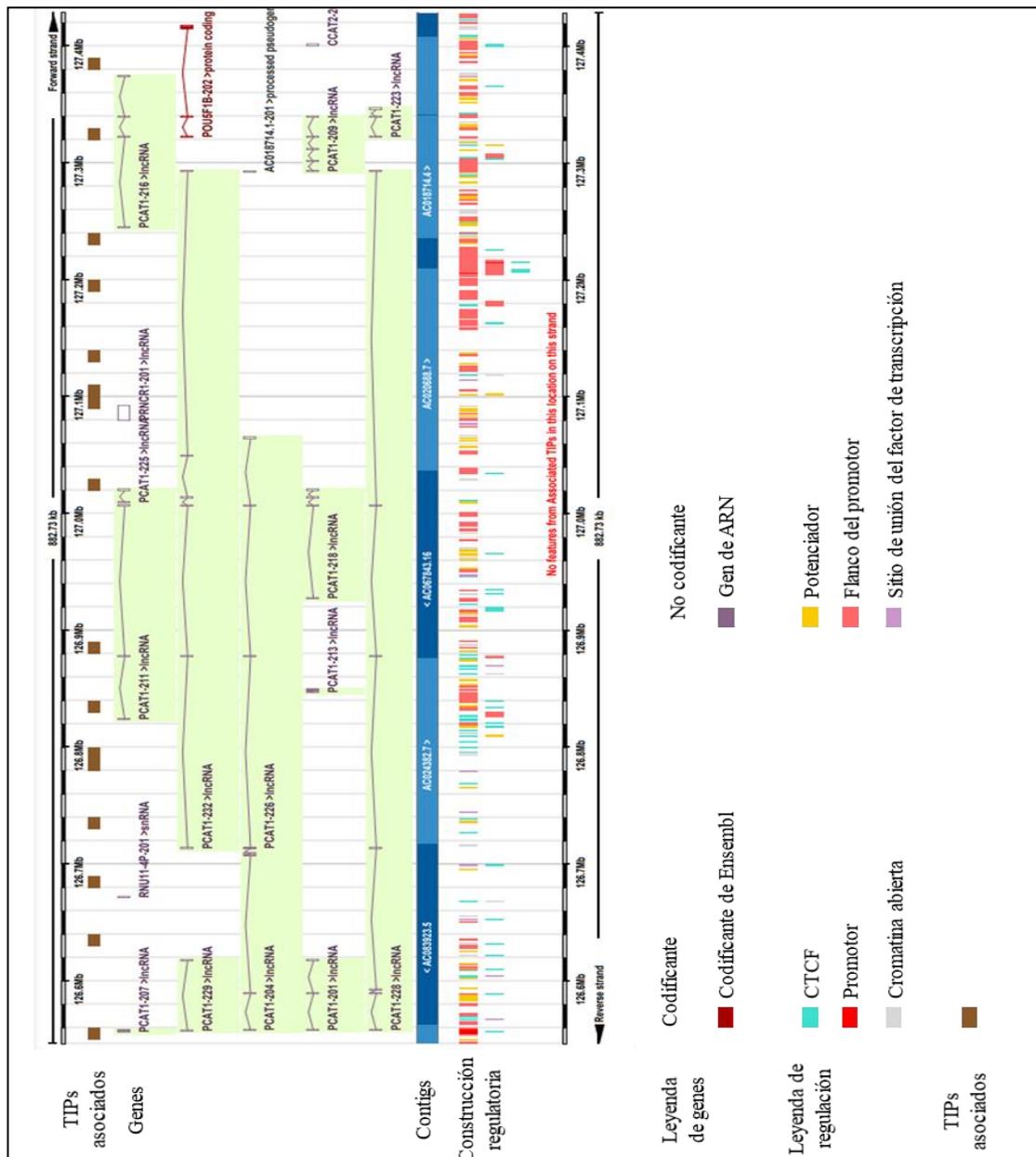
Navegador de genoma para el gen codificante MYC



Por último, se observa el navegador de genoma para el protooncogén *MYC* y su respectiva leyenda en la figura 11. Se puede observar un TIP a una distancia de aproximadamente 17 kb, no obstante, como se mencionó anteriormente, cabe destacar este gen debido a su relación con múltiples tipos de cáncer incluyendo la LMA.

Figura 12

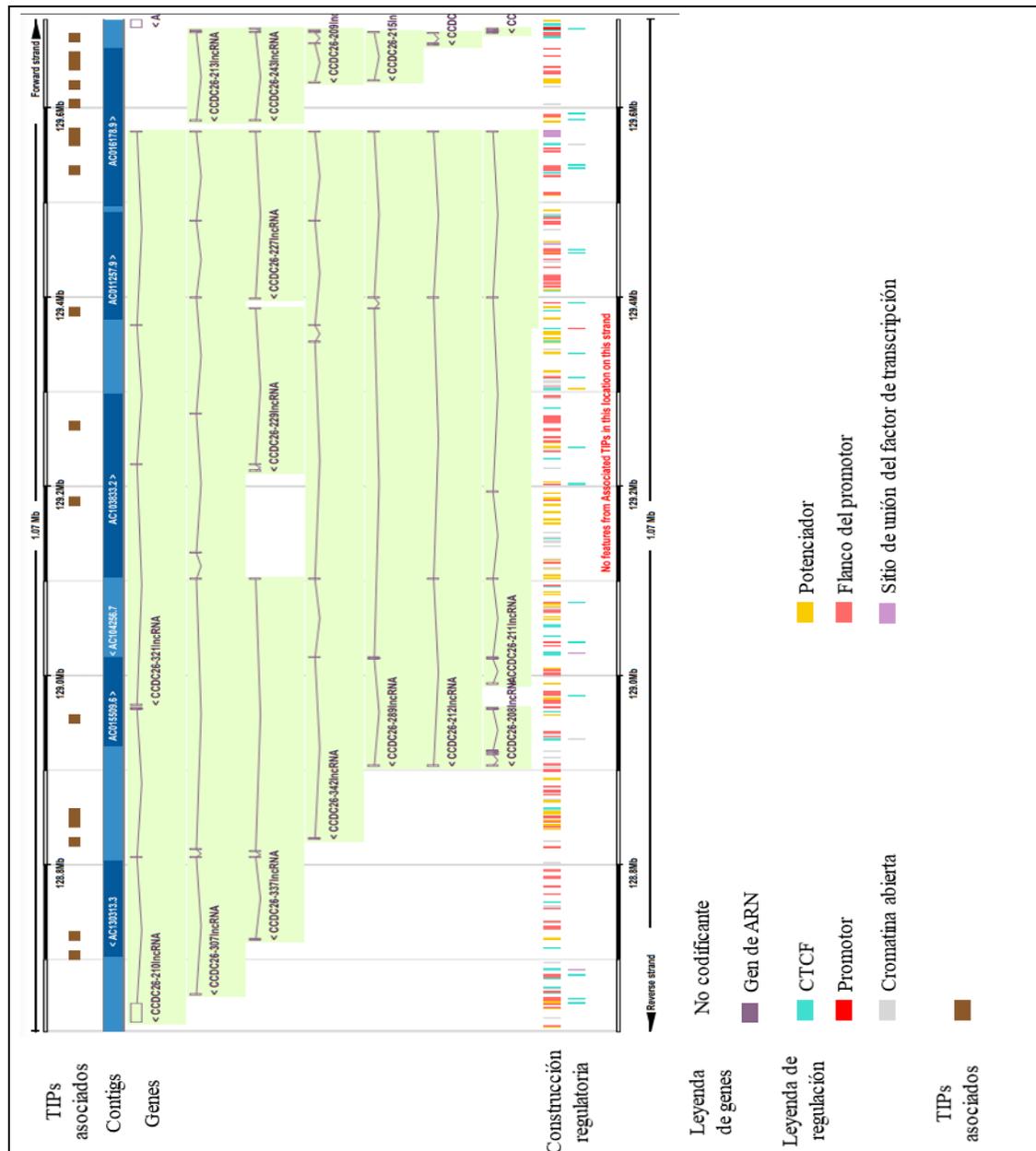
Navegador de genoma para el gen no codificante *PCAT1*



Por otro lado, la figura 12 muestra el navegador de genoma para el ARN largo no codificante (lncRNA) *PCAT1* y su respectiva leyenda. Además, muestra las posiciones de 14 TIPs de HERV-K asociados a LMA. Se puede observar que los TIPs superponen al gen en varias posiciones y además están ubicados sobre todas las regiones de interés del gen.

Figura 14

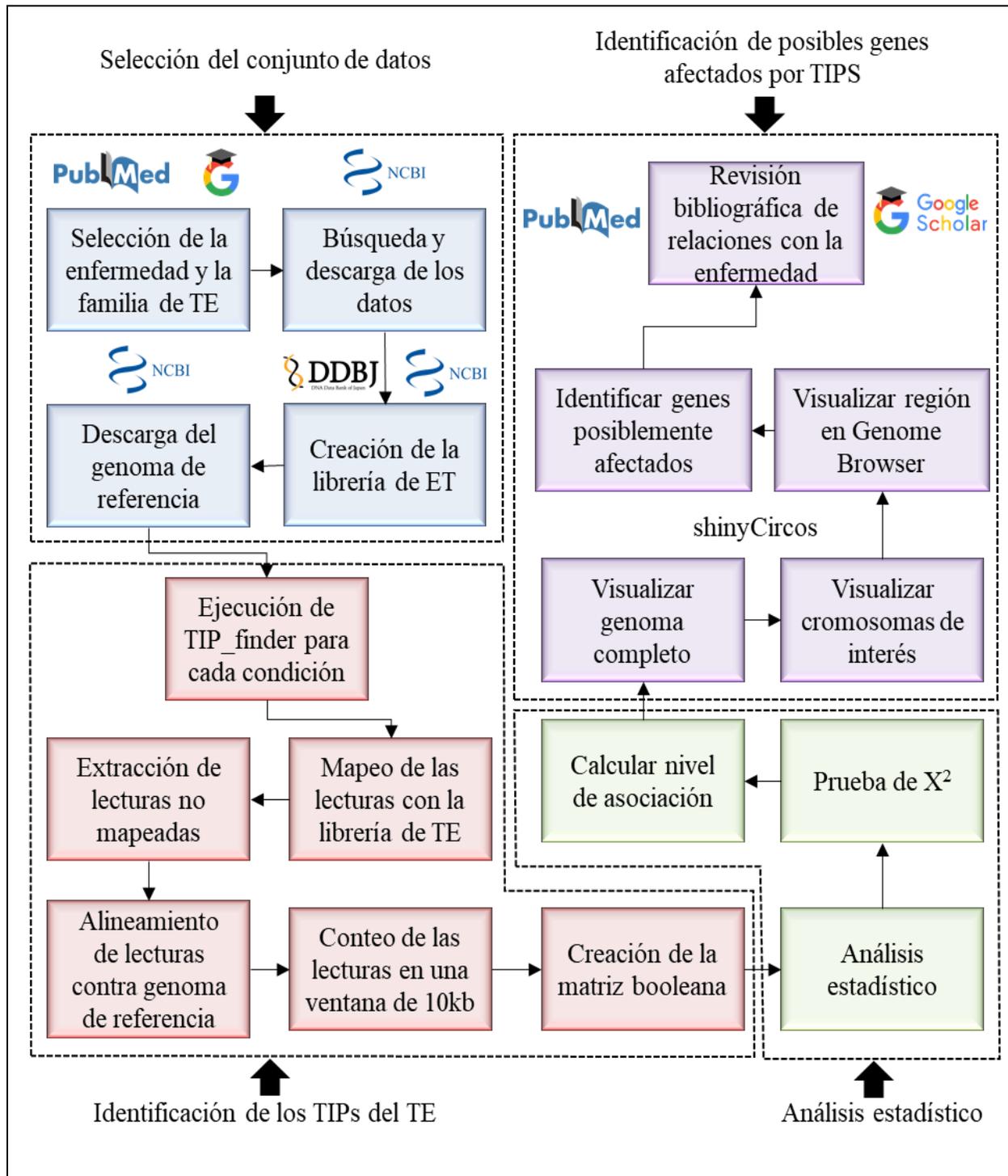
Navegador de genoma para el gen no codificante CCDC26



Por último, la figura 14 muestra el navegador de genoma para el lincRNA *CCDC26* y su respectiva leyenda. Además, muestra las posiciones de 14 TIPs de HERV-K asociados a LMA. Se puede observar que los TIPs superponen al gen en varias posiciones y además están ubicados sobre todas las regiones de interés del gen con excepción del promotor.

Figura 15

Metodología para la identificación de TIPS de una familia de elementos transponibles y su correlación con enfermedades específicas



Finalmente, se plantea la metodología para la replicación de este estudio, es decir, la identificación de TIPS de una familia de elementos transponibles, en este caso HERV-K y su correlación con una enfermedad específica, en este caso, leucemia mieloide aguda.

2.1.2.1 Discusión.

Los resultados evidenciaron asociación estadística entre TIPS de HERV-K y leucemia mieloide aguda, además, en todos los genes identificados en la región de interés se presentó al menos un TIP, con la excepción del gen *MYC*, lo que podría causar la desregulación presentada en los pacientes con la enfermedad. La tabla 2 muestra los diez TIPS de HERV-K identificados con mayor nivel de asociación a LMA, dando información como la posición, cantidad de pacientes/controles que presentaron este TIP y el porcentaje de asociación a la enfermedad. En total se encontraron 101 TIPS con un porcentaje de asociación mínimo del 44,4% y máximo del 56,6%. Estos TIPS en su mayoría (70) se encuentran ubicados en el cromosoma 8, razón por la cual se definió este cromosoma como el cromosoma de interés. De igual manera, en la figura 5 se pueden observar la cantidad de TIPS encontrados en cada cromosoma para pacientes y controles; los cromosomas 1, 7, 8, 9, 11, 12, 16, 18, 19 y X presentan una alta cantidad de TIPS, no obstante, el cromosoma 8 es el que presenta el nivel más elevado de TIPS en pacientes con LMA y no se presenta el mismo resultado en controles, reforzando los resultados encontrados en la tabla 2.

Por otro lado, para evitar sesgos relacionados a los metadatos, se realizó un análisis estadístico separando los datos de los pacientes enfermos en los dos grupos mencionados en la metodología, los cuales son pacientes con cariotipo normal y pacientes con alteraciones cromosómicas. Se encontró que los pacientes con cariotipo normal presentaron 489 TIPS asociados y los pacientes con alteraciones cromosómicas presentaron 224 TIPS asociados (sin

aplicar el filtro de p-valor menor a 1×10^{-5}). Además, se observó que los dos grupos compartían 154 TIPS de los cuales 103 se encontraban en la región de interés del cromosoma 8 y no se observaron agrupaciones relacionadas al tipo de cariotipo presentado. Esto valida los resultados presentados utilizando el grupo completo y por esta razón, se decidió presentar estos debido a la mayor fiabilidad presentada con respecto al P-valor. De igual manera, con respecto al género, al momento de escoger los datos se aseguró que incluyeran cantidades equivalentes de pacientes masculinos y femeninos, teniendo así, 33 pacientes masculinos y 33 pacientes femeninos para evitar agrupaciones relacionadas al género. Finalmente, con respecto a la edad también se aseguró que tanto los pacientes como los controles estuvieran distribuidos en un rango amplio de 20 a 90 años para que los resultados no se agruparan dependiendo de la edad.

Una vez identificado el cromosoma 8 como cromosoma de interés, se procedió a hacer un acercamiento (Figura 6) para identificar la región en donde se concentraban los TIPS asociados, siendo esta desde la base 125.000.000 hasta la 130.000.000. Al observar esta región en el navegador de genoma de Ensembl (Figura 7), se encontró que esta región corresponde a la banda del cromosoma 8q24, más específicamente 8q24.13-8q24.21. Al realizar una revisión de la literatura, se encontró que esta es una región de interés ya identificada para la leucemia mieloide aguda (Kamath et al., 2008; L'Abbate et al., 2018), en donde se han encontrado fusiones de genes entre el gen *NSMCE2* y ARNs largos no codificantes asociados a la enfermedad (Chinen et al., 2014), amplicones del protooncogén *MYC* debido a quimeras postranscripcionales y ARNs circulares (generalmente *PVT1*) en pacientes con cariotipo mutado (Ghetti et al., 2020; L'Abbate et al., 2018) y la patogénesis de la enfermedad asociada con la sobreexpresión de los genes *TRIB1* y *MYC* (Röthlisberger et al., 2007).

Estos estudios demuestran la relación de esta región con la leucemia mieloide aguda, por lo que se provee información sobre los genes codificantes y no codificantes de esta región como se muestra en la tabla 3 para entender más a fondo la relación de cada uno de estos genes con la enfermedad y como podrían verse afectados por los TIPS de HERV-K. Como se mencionó anteriormente, cabe destacar los genes codificantes *TRIB1*, *LRATD2*, *POU5F1B* y *MYC* y los genes no codificantes *PCAT1*, *PVT1* y *CCDC26*, por lo que se hace un análisis extensivo de estos genes, no obstante, los otros genes encontrados en esta región y afectados por TIPS se proponen para un posterior análisis (Apéndices G-I).

El homólogo 1 de Tribbles (Figura 8) es una proteína quinasa codificada por el gen *TRIB1*, la cual es fundamental como reguladora del ciclo celular, participando en la diferenciación, metabolismo, proliferación y estrés celular (Eyers et al., 2017). La sobreexpresión de *TRIB1* se ha visto asociada con la patogénesis de la leucemia mieloide aguda debido a que induce mayor fosforilación de la vía de señalización MAPK/ERK en las células hematopoyéticas lo que se ha visto estrechamente relacionado con la supresión de la apoptosis (Keeshan et al., 2016; Liang et al., 2013; Nakamura, 2015; Röthlisberger et al., 2007; Yokoyama et al., 2010). La figura 8 muestra la anotación de los transcritos de *TRIB1* con sus regiones de interés, en donde se puede observar un TIP de HERV-K en las regiones del promotor y el flanco del promotor. Tal inserción sugiere que el correcto funcionamiento de este gen podría verse afectado, debido a que el promotor es la región encargada de activar el inicio de la transcripción y una inserción en este podría causar la sobreexpresión de *TRIB1* en pacientes con LMA.

La sobreexpresión del gen *LRATD2* (*LRAT Domain Containing 2*, en inglés) (Figura 9) también llamado *FAM84B* se ha visto asociado con varios tipos de cáncer como carcinoma de células escamosas del esófago (Cheng et al., 2016), cáncer de próstata (Wong et al., 2017) y

gástrico (Y. Zhang et al., 2019). Con respecto a la LMA, *FAM84B* ha sido encontrado como un gen ubicado en la región de interés 8q24 asociada a LMA y se encuentra amplificado en pacientes con cariotipo mutado de LMA (Kamath et al., 2008; L'Abbate et al., 2018), no obstante, en la literatura no se encuentran estudios sobre el análisis de la expresión de este gen y su relación con la enfermedad. Por otro lado, la figura 9 muestra un TIP que superpone al gen y a las regiones del promotor y flanco del promotor. Este TIP podría ser una de las causas de la desregulación y funcionamiento anormal del gen. Por lo tanto, se propone este gen para un posterior análisis de su patrón de expresión en células hematopoyéticas mieloides debido a que se ha encontrado relación con otros tipos de cáncer.

El factor de transcripción de tipo 5 de dominio POU 1B (*POU5F1B*) (Figura 10) es un pseudogen que puede codificar una proteína similar al factor de transcripción POU clase 5 homeobox 1, la cual ha demostrado ser un activador transcripcional débil que puede desempeñar un papel en la carcinogénesis. Se ha reportado que la sobreexpresión de *POU5F1B* podría evitar la función de oncomirs (microARNs asociados al cáncer) dirigidos a genes esenciales para la represión celular y mediante la unión competitiva a los oncomirs podría inhibir la AML de alguna forma (Zhu et al., 2020), por consiguiente, se ha reportado una baja expresión de este gen en pacientes con AML, por lo que es usado como un factor de pronóstico de la enfermedad (Yi et al., 2019; Zhu et al., 2020). La figura 10 muestra las regiones de interés del gen, en donde se pueden observar dos TIPs de HERV-K que superponen al gen en diferentes posiciones. Estas inserciones podrían fragmentar al gen, afectando su correcto funcionamiento y causando la baja expresión de este presentada en pacientes con LMA.

Por último, el gen *MYC* (Figura 11), es un protooncogén que codifica una fosfoproteína que desempeña un papel en la progresión del ciclo celular, proliferación, hematopoyesis,

apoptosis y diferenciación celular. No obstante, al ser un protooncogén, su desregulación, expresión anormal y mutación se ha visto asociada con múltiples tipos de cáncer (Dang, 2012; Stine et al., 2015), incluyendo la leucemia mieloide aguda, en donde la sobreexpresión de este gen se puede usar como un factor de diagnóstico, perfil de riesgo citogenético y supervivencia promedio (Mughal et al., 2017; Ohanian et al., 2019; Yun et al., 2019). A diferencia de otros genes mostrados en este estudio, el gen *MYC* no presenta TIPS dentro de sus regiones de interés cómo se observa en la figura 11, sino que presenta un TIP a aproximadamente 17 kb. No obstante, a pesar de que se desconoce la distancia en la que un elemento transponible puede afectar a un gen o sus vías metabólicas, vale la pena analizar esta relación en un estudio futuro para evaluar la posible interacción entre el gen y el TIP de HERV-K y los potenciales impactos sobre su expresión.

Por otro lado, los ARNs largos no codificantes (lncRNA, por sus siglas en inglés) son moléculas de ARN funcional de más de 200 nucleótidos que no son traducidas a proteínas, cuya función se ha visto asociada con regulación de la transcripción de genes, regulación post transcripcional y regulación epigenética (Fernandes et al., 2019). Además, la mutación y desregulación de estos se ha visto implicada en la patogénesis y desarrollo de varios tipos de cáncer como cáncer de próstata, cáncer de pulmón, cáncer gástrico, cáncer colorrectal y leucemia mieloide aguda, entre otros (Bhan et al., 2017; Huarte, 2015). Por esta razón, este estudio evaluó los ARNs largos no codificantes que han presentado asociación con la LMA, tales como: *PCAT1*, *PVT1* y *CCDC26*.

El transcrito asociado al cáncer de próstata (*PCAT1*) (Figura 12) es un ARN largo no codificante que promueve la proliferación celular y un represor de la transcripción que desregula la expresión de la proteína BRCA2 que actúa como un supresor de tumores. Por esta razón, la

sobreexpresión de *PCATI* se ha visto asociada con diferentes tipos de cáncer tales como cáncer de próstata, cáncer colorrectal, cáncer gástrico, cáncer de pulmón y leucemia mieloide aguda, entre otros (Ghafouri-Fard et al., 2020). Con respecto a la leucemia mieloide aguda, se ha demostrado que la expresión de *PCATI* es más alta en pacientes de LMA que en controles y que interactúa con FZD6 activando la cascada de señalización Wnt/beta-catenina lo que podría desempeñar un papel en la patogénesis de la LMA (Yuan et al., 2019). La figura 12 muestra algunos de los transcritos de *PCATI* con sus regiones de interés, además, se pueden observar 14 TIPS de HERV-K superponiendo varios de los transcritos y todas sus regiones de interés. Inserciones en el potenciador del gen podrían sugerir un aumento de la expresión del gen mencionado, lo cual concuerda con la mayor expresión observada en pacientes con AML.

El ARN largo no codificante *PVT1* (*Plasmacytoma Variant Translocation 1*) (Figura 13) es un candidato a oncogén debido a que la sobreexpresión de este se ha visto asociada con la patogénesis de múltiples tipos de cáncer. Este gen se puede presentar en dos isoformas, como ARN circular y como ARN linear no codificante, las cuales se han visto asociadas con malignidades hematológicas, alteraciones genómicas, progresión de enfermedades y regulación de la respuesta del sistema inmune (Ghetti et al., 2020). De igual manera, la sobreexpresión, multiplicación del número de sus copias y participación en transcripciones quiméricas se han visto asociadas con la leucemia mieloide aguda, aunque, esto se presenta generalmente en pacientes con cariotipo anormal (Dieter et al., 2020; Ghetti et al., 2020; Izadifard et al., 2018; L'Abbate et al., 2018). En la figura 13 se observan algunos transcritos de *PVT1* y siete TIPS de HERV-K superponiendo a este y a todas sus regiones de interés con la excepción del sitio de unión del factor de transcripción. Además, en varios transcritos se pueden ver hasta seis TIPS insertados en diferentes posiciones. Estas inserciones potencialmente causan la fragmentación

del gen o afectan sus regiones de interés. Esta fragmentación podría ser la causa de la fusión con otros genes y la inserción en regiones como el promotor y potenciador podrían causar la sobreexpresión del gen mostrada en pacientes con LMA.

Finalmente, el gen *CCDC26* (Figura 14) es un ARN largo no codificante cuya función es desconocida actualmente, no obstante, se ha asociado con la carcinogénesis de múltiples tipos de cáncer como gliomas, cáncer gastrointestinal, cáncer de páncreas y leucemia mieloide aguda (Peng & Jiang, 2016; S. Wang et al., 2018; Yan et al., 2019). *CCDC26* ha sido usado como biomarcador para la predicción de LMA debido a que presenta una mayor expresión en pacientes con la enfermedad (C. Chen et al., 2019; Izadifard et al., 2018), además, se cree que este gen controla el crecimiento de las células cancerígenas al causar la sobreexpresión de KIT, una tirosina quinasa asociada a la LMA (Hirano et al., 2015). La figura 14 muestra algunos transcritos de *CCDC26* con sus respectivas regiones de interés y 14 TIPS de HERV-K superponiendo al gen en varias ubicaciones y a todas sus regiones con excepción del promotor. Si se observa el segundo transcrito (*CCDC26-307*) se puede notar que este transcrito está superpuesto por 10 TIPS, lo que sugiere la fragmentación de este gen o la alteración de alguna de sus regiones, lo cual podría ser causante del mal funcionamiento gen y la sobreexpresión de este.

Todos los genes analizados en este estudio han sido reportados como genes asociados con la leucemia mieloide aguda, con la excepción de *LRATD2*. Además, la mayoría de los TIPS identificados se encuentran ubicados en la banda del cromosoma 8q24, una región de interés previamente identificada por otros estudios (Chinen et al., 2014; L'Abbate et al., 2018).

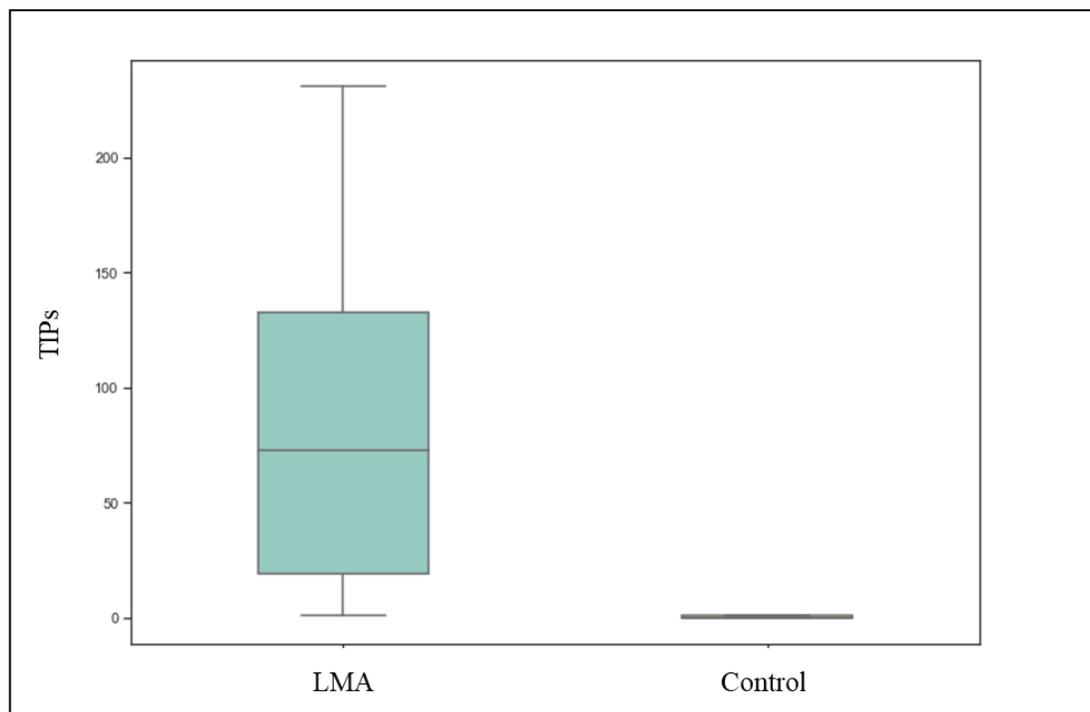
Por otro lado, como se puede apreciar en las figuras 8 a 14, todos los genes codificantes y no codificantes presentan al menos un TIP de HERV-K insertado o superponiendo alguna de sus regiones de interés, con la excepción del protooncogén *MYC*, lo cual sugiere una posible relación

con la desregulación de estos genes en la leucemia mieloide aguda. Esto concuerda con los resultados encontrados en la literatura, dado que inserciones de elementos transponibles se han visto asociadas con cambios en la expresión de los genes y en la regulación epigenética de los mismos (Slotkin & Martienssen, 2007; Zeng et al., 2018).

No obstante, no se puede conocer específicamente la forma en la que el TIP afecta al gen, ya sea fragmentación o desplazamiento de este o de sus regiones de interés. Por lo tanto, se sugiere realizar estudios futuros para identificar específicamente la forma en la que el TIP está afectando a cada gen y si esto es el causante del mal funcionamiento y desregulación de este. Por otro lado, para el caso del gen *LRATD2*, como se mencionó anteriormente, se necesitan estudios de análisis del patrón de expresión en células hematopoyéticas mieloides de pacientes con LMA y controles para evaluar su posible relación con la enfermedad.

Figura 16

Cantidad de TIPS de HERV-K asociados a LMA por condición



Los resultados presentados en este estudio permiten proponer las bases para la creación un nuevo método de predicción y diagnóstico de leucemia mieloide aguda basado en TIPS de HERV-K usando el genoma completo o de manera más rápida usando el cromosoma 8 (esto debido a la cantidad de TIPS identificados en este cromosoma). Como se observa en la figura 16, la cantidad promedio de TIPS de HERV-K presentados por paciente de LMA es de 80 TIPS aproximadamente, mientras que en los controles no hay presencia de estos. Por lo tanto, se sugiere que estos TIPS de HERV-K se encuentran asociados con la LMA y pueden ser usados para el diagnóstico y predicción de esta. No obstante, se requieren estudios futuros usando una mayor cantidad de pacientes para confirmar estos resultados.

Estas observaciones pueden interpretarse de tres maneras: los TIPS de HERV-K son la causa directa de la patogénesis de la leucemia mieloide aguda, los TIPS son una consecuencia de los mecanismos causantes de la LMA o los TIPS son una causa asociada a otro mecanismo y juntos son los causantes de la enfermedad. Sin embargo, para entender mejor el impacto de los TIPS en la enfermedad son necesarios otros datos, como los patrones de expresión de genes en la región de interés en pacientes y controles y el estado de metilación de ADN en la región. Al confirmar este impacto se puede establecer si estos resultados pueden ser utilizados como método de predicción y diagnóstico, en el caso de que los TIPS sean la causa de la enfermedad, o solo como método de diagnóstico, en el caso de que los TIPS sean una consecuencia de la enfermedad. Por otro lado, también es importante evaluar el estado de los TIPS de HERV-K en pacientes con leucemia mieloide aguda en sus diferentes etapas, para conocer más a fondo este mecanismo y proveer un método de diagnóstico y predicción más acertada debido a que es una enfermedad que se desarrolla rápidamente.

Finalmente, la metodología propuesta en la figura 15 provee un conjunto de acciones generalizado y replicable para analizar la correlación entre TIPS de diferentes familias de elementos transponibles y alguna enfermedad específica. Esto con el objetivo de que a futuro se propongan nuevos métodos de predicción y diagnóstico de diferentes enfermedades usando el genoma completo.

3. Conclusiones

Como se mencionó anteriormente, la leucemia mieloide aguda es un problema a nivel mundial debido a que es una enfermedad con una tasa de supervivencia de menos del 30%, por lo tanto, es importante contribuir a la creación de nuevas estrategias de diagnóstico, predicción y tratamiento de la misma, para así, mejorar la calidad de vida de los pacientes. Por esta razón, se concluye que los TIPS de HERV-K encontrados en este estudio y más específicamente los encontrados en la región q24.13 y q24.21 del cromosoma 8 (125.000.000-130.000.000) se encuentran asociados estadísticamente con la leucemia mieloide aguda en un porcentaje entre el 44,4 al 56,6% y con aproximado de 80 TIPS por paciente con la enfermedad. Este método podría utilizarse como método de diagnóstico y/o predicción por medio de una secuenciación completa del genoma del paciente o una secuenciación dirigida de la región de interés del cromosoma 8 y a continuación, utilizar la metodología propuesta en este proyecto para identificar la cantidad de TIPS que presenta el paciente y así hacer una estimación de la probabilidad de que la persona pueda tener leucemia mieloide aguda. No obstante, para poder utilizar este método en el área de la medicina es importante realizar los estudios mencionados anteriormente para que esta predicción sea más acertada.

Por otro lado, también se concluye que estos TIPS podrían afectar el correcto funcionamiento de los genes *TRIB1*, *LRATD2*, *POU5F1B*, *MYC*, *PCAT1*, *PVT1* y *CCDC26* al alterar su expresión y/o regulación epigenética, lo cual se ha visto asociado con la patogénesis de la LMA, sin embargo, se requiere conocer la forma específica en la que los TIPS afectan a estos genes. Finalmente, la metodología de este proyecto puede aplicarse utilizando cualquier familia de elemento transponible y una enfermedad asociada, lo cual podría permitir la creación de nuevos métodos de diagnóstico y/o predicción de diferentes enfermedades basada en TIPS.

Referencias Bibliográficas

- AlMutairi, M., Parine, N. R., Shaik, J. P., Aldhaian, S., Azzam, N. A., Aljebreen, A. M., Alharbi, O., Almadi, M. A., Al-Balbeesi, A. O., & Alanazi, M. (2019). Association between polymorphisms in PRNCR1 and risk of colorectal cancer in the Saudi population. *PloS One*, *14*(9), e0220931. <https://doi.org/10.1371/journal.pone.0220931>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Barrón, M. G., Fiston-Lavier, A.-S., Petrov, D. A., & González, J. (2014). Population Genomics of Transposable Elements in *Drosophila*. *Annual Review of Genetics*, *48*(1), 561–581. <https://doi.org/10.1146/annurev-genet-120213-092359>
- Bergallo, M., Montanari, P., Mareschi, K., Merlino, C., Berger, M., Bini, I., Daprà, V., Galliano, I., & Fagioli, F. (2017). Expression of the pol gene of human endogenous retroviruses HERV-K and -W in leukemia patients. *Archives of Virology*, *162*(12), 3639–3644. <https://doi.org/10.1007/s00705-017-3526-7>
- Bhan, A., Soleimani, M., & Mandal, S. S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Research*, *77*(15), 3965–3981. <https://doi.org/10.1158/0008-5472.CAN-16-2634>
- Bradley-Golem, S. (2013). *Human Endogenous Retrovirus (HERV) Insertional Polymorphisms*. Department of Biological Sciences.
- Brodsky, I., Foley, B., & Gillespie, D. (1993). Expression of human endogenous retrovirus (HERV-k) in chronic myeloid leukemia. *Leukemia and Lymphoma*, *11*(s1), 119–123.

<https://doi.org/10.3109/10428199309047874>

- Brown, D. N., Caffa, I., Cirmena, G., Piras, D., Garuti, A., Gallo, M., Alberti, S., Nencioni, A., Ballestrero, A., & Zoppoli, G. (2016). Squalene epoxidase is a bona fide oncogene by amplification with clinical relevance in breast cancer. *Scientific Reports*, *6*(1), 19435. <https://doi.org/10.1038/srep19435>
- Burghaus, S., Fasching, P. A., Häberle, L., Rübner, M., Büchner, K., Blum, S., Engel, A., Ekici, A. B., Hartmann, A., Hein, A., Beckmann, M. W., & Renner, S. P. (2017). Genetic risk factors for ovarian cancer and their role for endometriosis risk. *Gynecologic Oncology*, *145*(1), 142–147. <https://doi.org/https://doi.org/10.1016/j.ygyno.2017.02.022>
- Büscher, K., Hahn, S., Hofmann, M., Trefzer, U., Özel, M., Sterry, W., Löwer, J., Löwer, R., Kurth, R., & Denner, J. (2006). Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines. *Melanoma Research*, *16*(3), 223–234. <https://doi.org/10.1097/01.cmr.0000215031.07941.ca>
- Carpentier, M.-C., Manfroi, E., Wei, F.-J., Wu, H.-P., Lasserre, E., Llauro, C., Debladis, E., Akakpo, R., Hsing, Y.-I., & Panaud, O. (2019). Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nature Communications*, *10*(1), 1–24. <https://doi.org/10.1038/s41467-018-07974-5>
- Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology*, *22*(6), 1503–1517. <https://doi.org/10.1111/mec.12170>
- Cegolon, L., Salata, C., Weiderpass, E., Vineis, P., Palù, G., & Mastrangelo, G. (2013). Human endogenous retroviruses and cancer prevention: evidence and prospects. *BMC Cancer*, *13*(1), 1–4. <https://doi.org/10.1186/1471-2407-13-4>
- Chande, A. T., Wang, L., Rishishwar, L., Conley, A. B., Norris, E. T., Valderrama-Aguirre, A.,

- & Jordan, I. K. (2018). Global Distribution of Genetic Traits (GADGET) web server: polygenic trait scores worldwide. *Nucleic Acids Research*, *46*(1), 121–126. <https://doi.org/10.1093/nar/gky415>
- Chen, C., Wang, P., Mo, W., Zhang, Y., Zhou, W., Deng, T., Zhou, M., Chen, X., Wang, S., & Wang, C. (2019). lncRNA-CCDC26, as a novel biomarker, predicts prognosis in acute myeloid leukemia. *Oncology Letters*, *18*(3), 2203–2211. <https://doi.org/10.3892/ol.2019.10591>
- Chen, J., Wu, J. S., Mize, T., Moreno, M., Hamid, M., Servin, F., Bashy, B., Zhao, Z., Jia, P., Tsuang, M. T., Kendler, K. S., Xiong, M., & Chen, X. (2019). A Frameshift Variant in the CHST9 Gene Identified by Family-Based Whole Genome Sequencing Is Associated with Schizophrenia in Chinese Population. *Scientific Reports*, *9*(1), 1–9. <https://doi.org/10.1038/s41598-019-49052-w>
- Chen, S., Chen, Y., Zhu, Z., Tan, H., Lu, J., Qin, P., & Xu, L. (2020). Identification of the key genes and microRNAs in adult acute myeloid leukemia with FLT3 mutation by bioinformatics analysis. *International Journal of Medical Sciences*, *17*(9), 1269–1280. <https://doi.org/10.7150/ijms.46441>
- Chen, X., Li, D., & Birol, I. (2019). ERVcaller: Identifying polymorphic endogenous retrovirus and other transposable element insertions using whole-genome sequencing data. *Bioinformatics*, *35*(20), 3913–3922. <https://doi.org/10.1093/bioinformatics/btz205>
- Cheng, C., Cui, H., Zhang, L., Jia, Z., Song, B., Wang, F., Li, Y., Liu, J., Kong, P., Shi, R., Bi, Y., Yang, B., Wang, J., Zhao, Z., Zhang, Y., Hu, X., Yang, J., He, C., Zhao, Z., ... Cui, Y. (2016). Genomic analyses reveal FAM84B and the NOTCH pathway are associated with the progression of esophageal squamous cell carcinoma. *GigaScience*, *5*(1), 1–14.

<https://doi.org/10.1186/s13742-015-0107-0>

- Chinen, Y., Sakamoto, N., Nagoshi, H., Taki, T., Maegawa, S., Tatekawa, S., Tsukamoto, T., Mizutani, S., Shimura, Y., Yamamoto-Sugitani, M., Kobayashi, T., Matsumoto, Y., Horiike, S., Kuroda, J., & Taniwaki, M. (2014). 8q24 amplified segments involve novel fusion genes between NSMCE2 and long noncoding RNAs in acute myelogenous leukemia. *Journal of Hematology & Oncology*, 7, 68. <https://doi.org/10.1186/s13045-014-0068-2>
- Cochran, W. (1954). Some Methods for Strengthening the Common X^2 Tests. *International Biometric Society*, 10, 417–451. <http://www.jstor.org/stable/3001616>
- Colombo, A. R., Triche, T., & Ramsingh, G. (2018). Transposable Element Expression in Acute Myeloid Leukemia Transcriptome and Prognosis. *Scientific Reports*, 8(1), 16449. <https://doi.org/10.1038/s41598-018-34189-x>
- Coordinators, N. R. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 44(D1), D7–D19. <https://doi.org/10.1093/nar/gkv1290>
- Dalcin, L. D., Paz, R. R., Kler, P. A., & Cosimo, A. (2011). Parallel distributed computing using Python. *Advances in Water Resources*, 34(9), 1124–1139. <https://doi.org/https://doi.org/10.1016/j.advwatres.2011.04.013>
- Dang, C. V. (2012). MYC on the path to cancer. *Cell*, 149(1), 22–35. <https://doi.org/10.1016/j.cell.2012.03.003>
- Davis, A. S., Viera, A. J., & Mead, M. D. (2014). Leukemia: An overview for primary care. *American Family Physician*, 89(9), 731–738.
- Denne, M., Sauter, M., Armbruester, V., Licht, J. D., Roemer, K., & Mueller-Lantsch, N. (2007). Physical and Functional Interactions of Human Endogenous Retrovirus Proteins

- Np9 and Rec with the Promyelocytic Leukemia Zinc Finger Protein. *Journal of Virology*, 81(11), 5607–5616. <https://doi.org/10.1128/jvi.02771-06>
- Depil, S., Roche, C., Dussart, P., & Prin, L. (2002). Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. *Leukemia*, 16(2), 254–259. <https://doi.org/10.1038/sj.leu.2402355>
- Dieter, C., Lourenco, E. D., & Lemos, N. E. (2020). Association of long non-coding RNA and leukemia: A systematic review. *Gene*, 735, 144405. <https://doi.org/10.1016/j.gene.2020.144405>
- Ding, W., Yang, H., Gong, S., Shi, W., Xiao, J., Gu, J., Wang, Y., & He, B. (2017). Candidate miRNAs and pathogenesis investigation for hepatocellular carcinoma based on bioinformatics analysis. *Oncology Letters*, 13(5), 3409–3414. <https://doi.org/10.3892/ol.2017.5913>
- Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015). Acute Myeloid Leukemia. *The New England Journal of Medicine*, 373(12), 1136–1152. <https://doi.org/10.1056/NEJMra1406184>
- Dolci, M., Favero, C., Toumi, W., Favi, E., Tarantini, L., Signorini, L., Basile, G., Bollati, V., D'Alessandro, S., Bagnoli, P., Ferrante, P., & Delbue, S. (2020). Human Endogenous Retroviruses Long Terminal Repeat Methylation, Transcription, and Protein Expression in Human Colon Cancer. *Frontiers in Oncology*, 10(October), 1–10. <https://doi.org/10.3389/fonc.2020.569015>
- Drongitis, D., Aniello, F., Fucci, L., & Donizetti, A. (2019). Roles of Transposable Elements in the Different Layers of Gene Expression Regulation. *International Journal of Molecular Sciences*, 20(22), 5755. <https://doi.org/10.3390/ijms20225755>

- Edwards, D, Stajich, J., & Hansen, D. (2009). *Bioinformatics* (David Edwards, J. Stajich, & D. Hansen (eds.); 1st ed.). Springer New York. <https://doi.org/10.1007/978-0-387-92738-1>
- El-Khazragy, N., Elayat, W., Matbouly, S., Seliman, S., Sami, A., Safwat, G., & Diab, A. (2019). The prognostic significance of the long non-coding RNAs “CCAT1, PVT1” in t(8;21) associated Acute Myeloid Leukemia. *Gene*, 707, 172–177. <https://doi.org/10.1016/j.gene.2019.03.055>
- Ewing, A. D. (2015). Transposable element detection from whole genome sequence data. *Mobile DNA*, 6(1), 24. <https://doi.org/10.1186/s13100-015-0055-3>
- Eyers, P. A., Keeshan, K., & Kannan, N. (2017). Tribbles in the 21st Century: The Evolving Roles of Tribbles Pseudokinases in Biology and Disease. *Trends in Cell Biology*, 27(4), 284–298. <https://doi.org/10.1016/j.tcb.2016.11.002>
- Feng, Y., Shen, Y., Chen, H., Wang, X., Zhang, R., Peng, Y., Lei, X., Liu, T., Liu, J., Gu, L., Wang, F., Yang, Y., Bai, J., Wang, J., Zhao, W., & He, A. (2018). Expression profile analysis of long non-coding RNA in acute myeloid leukemia by microarray and bioinformatics. *Cancer Science*, 109(2), 340–353. <https://doi.org/10.1111/cas.13465>
- Fernandes, J., Acuña, S., Aoki, J., Floeter-Winter, L., & Muxel, S. (2019). Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Non-Coding RNA*, 5(1), 17. <https://doi.org/10.3390/ncrna5010017>
- Garcia-Montojo, M., Doucet-O’Hare, T., Henderson, L., & Nath, A. (2018). Human endogenous retrovirus-K (HML-2): a comprehensive review. *Critical Reviews in Microbiology*, 44(6), 715–738. <https://doi.org/10.1080/1040841X.2018.1501345>
- Ge, H., Zhao, Y., Shi, X., Tan, Z., Chi, X., He, M., Jiang, G., Ji, L., & Li, H. (2019). Squalene epoxidase promotes the proliferation and metastasis of lung squamous cell carcinoma cells

- though extracellular signal-regulated kinase signaling. *Thoracic Cancer*, *10*(3), 428–436. <https://doi.org/10.1111/1759-7714.12944>
- Ghafouri-Fard, S., Dashti, S., & Taheri, M. (2020). PCAT1: An oncogenic lncRNA in diverse cancers and a putative therapeutic target. *Experimental and Molecular Pathology*, *114*, 104429. <https://doi.org/10.1016/j.yexmp.2020.104429>
- Ghetti, M., Vannini, I., Storlazzi, C. T., Martinelli, G., & Simonetti, G. (2020). Linear and circular PVT1 in hematological malignancies and immune response: two faces of the same coin. *Molecular Cancer*, *19*(1), 69. <https://doi.org/10.1186/s12943-020-01187-5>
- Guo, S., Li, B., Xu, X., Wang, W., Wang, S., Lv, T., & Wang, H. (2020). Construction of a 14-lncRNA risk score system predicting survival of children with acute myelocytic leukemia. *Experimental and Therapeutic Medicine*, *20*(2), 1521–1531. <https://doi.org/10.3892/etm.2020.8846>
- Hashemi, M., Bahari, G., Naderi, M., Sadeghi Bojd, S., & Taheri, M. (2016). Association of lnc-LAMC2-1:1 rs2147578 and CASC8 rs10505477 Polymorphisms with Risk of Childhood Acute Lymphoblastic Leukemia. *Asian Pacific Journal of Cancer Prevention: APJCP*, *17*(11), 4985–4989. <https://doi.org/10.22034/APJCP.2016.17.11.4985>
- Hayashi, H., Arao, T., Togashi, Y., Kato, H., Fujita, Y., De Velasco, M. A., Kimura, H., Matsumoto, K., Tanaka, K., Okamoto, I., Ito, A., Yamada, Y., Nakagawa, K., & Nishio, K. (2015). The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene*, *34*(2), 199–208. <https://doi.org/10.1038/onc.2013.547>
- He, Z., Zhu, J., Mo, J., Zhao, H., & Chen, Q. (2020). HBV DNA integrates into upregulated ZBTB20 in patients with hepatocellular carcinoma. *Molecular Medicine Reports*, *22*(1),

380–386. <https://doi.org/10.3892/mmr.2020.11074>

- Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., & Meyerson, M. (2014). Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Research*, *24*(7), 1053–1063. <https://doi.org/10.1101/gr.163659.113>
- Hénaff, E., Zapata, L., Casacuberta, J. M., & Ossowski, S. (2015). Jitterbug: Somatic and germline transposon insertion detection at single-nucleotide resolution. *BMC Genomics*, *16*(1), 1–16. <https://doi.org/10.1186/s12864-015-1975-5>
- Hirano, T., Yoshikawa, R., Harada, H., Harada, Y., Ishida, A., & Yamazaki, T. (2015). Long noncoding RNA, CCDC26, controls myeloid leukemia cell growth through regulation of KIT expression. *Molecular Cancer*, *14*(1), 90. <https://doi.org/10.1186/s12943-015-0364-7>
- Hohn, O., Hanke, K., & Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Frontiers in Oncology*, *3*, 246. <https://doi.org/10.3389/fonc.2013.00246>
- Hong, J. H., Jin, E.-H., Kang, H., Chang, I. A., Lee, S.-I., & Sung, J. K. (2019). Correlations between Genetic Polymorphisms in Long Non-Coding RNA PRNCR1 and Gastric Cancer Risk in a Korean Population. *International Journal of Molecular Sciences*, *20*(13), 3355. <https://doi.org/10.3390/ijms20133355>
- Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E., & Cronin, K. (2020). *SEER Cancer Statistics Review, 1975-2017*. National Cancer Institute. https://seer.cancer.gov/csr/1975_2017/
- Hu, L., Chen, S.-H., Lv, Q.-L., Sun, B., Qu, Q., Qin, C.-Z., Fan, L., Guo, Y., Cheng, L., & Zhou, H.-H. (2016). Clinical Significance of Long Non-Coding RNA CASC8 rs10505477 Polymorphism in Lung Cancer Susceptibility, Platinum-Based Chemotherapy Response,

- and Toxicity. *International Journal of Environmental Research and Public Health*, *13*(6), 545. <https://doi.org/10.3390/ijerph13060545>
- Hu, R., Zhong, P., Xiong, L., & Duan, L. (2017). Long Noncoding RNA Cancer Susceptibility Candidate 8 Suppresses the Proliferation of Bladder Cancer Cells via Regulating Glycolysis. *DNA and Cell Biology*, *36*(9), 767–774. <https://doi.org/10.1089/dna.2017.3785>
- Huang, R., Liao, X., & Li, Q. (2017). Identification of key pathways and genes in TP53 mutation acute myeloid leukemia: evidence from bioinformatics analysis. *Oncotargets and Therapy*, *11*, 163–173. <https://doi.org/10.2147/OTT.S156003>
- Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature Medicine*, *21*(11), 1253–1261. <https://doi.org/10.1038/nm.3981>
- Hughes, J. F., & Coffin, J. M. (2004). Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1668–1672. <https://doi.org/10.1073/pnas.0307885100>
- Izadifard, M., Pashaiefar, H., Yaghmaie, M., Montazeri, M., Sadraie, M., Momeny, M., Jalili, M., Ahmadvand, M., Ghaffari, S. H., Mohammadi, S., Alimoghaddam, K., & Ghavamzadeh, A. (2018). Expression Analysis of PVT1, CCDC26, and CCAT1 Long Noncoding RNAs in Acute Myeloid Leukemia Patients. *Genetic Testing and Molecular Biomarkers*, *22*(10), 593–598. <https://doi.org/10.1089/gtmb.2018.0143>
- Januszkiewicz-Lewandowska, D., Nowicka, K., Rembowska, J., Fichna, M., Zurawek, M., Derwich, K., & Nowak, J. (2013). Env gene expression of human endogenous retrovirus-k and human endogenous retrovirus-W in childhood acute leukemia cells. *Acta Haematologica*, *129*(4), 232–237. <https://doi.org/10.1159/000345407>

- Kamath, A., Tara, H., Xiang, B., Bajaj, R., He, W., & Li, P. (2008). Double-minute MYC amplification and deletion of MTAP, CDKN2A, CDKN2B, and ELAVL2 in an acute myeloid leukemia characterized by oligonucleotide-array comparative genomic hybridization. *Cancer Genetics and Cytogenetics*, *183*(2), 117–120. <https://doi.org/10.1016/j.cancergencyto.2008.02.011>
- Kastler, S., Honold, L., Luedeke, M., Kuefer, R., Möller, P., Hoegel, J., Vogel, W., Maier, C., & Assum, G. (2010). POU5F1P1, a putative cancer susceptibility gene, is overexpressed in prostatic carcinoma. *The Prostate*, *70*(6), 666–674. <https://doi.org/10.1002/pros.21100>
- Keeshan, K., Vieugué, P., Chaudhury, S., Rishi, L., Gaillard, C., Liang, L., Garcia, E., Nakamura, T., Omidvar, N., & Kogan, S. C. (2016). Co-operative leukemogenesis in acute myeloid leukemia and acute promyelocytic leukemia reveals C/EBP α as a common target of TRIB1 and PML/RARA. *Haematologica*, *101*(10), 1228–1236. <https://doi.org/10.3324/haematol.2015.138503>
- Kleiman, A., Senyuta, N., Tryakin, A., Sauter, M., Karseladze, A., Tjulandin, S., Gurtsevitch, V., & Mueller-Lantzsch, N. (2004). HERV-K(HML-2) GAG/ENV antibodies as indicator for therapy effect in patients with germ cell tumors. *International Journal of Cancer*, *110*(3), 459–461. <https://doi.org/10.1002/ijc.11649>
- L'Abbate, A., Tolomeo, D., Cifola, I., Severgnini, M., Turchiano, A., Augello, B., Squeo, G., D'Addabbo, P., Traversa, D., Daniele, G., Lonoce, A., Pafundi, M., Carella, M., Palumbo, O., Dolnik, A., Muehlematter, D., Schoumans, J., Van Roy, N., De Bellis, G., ... Storlazzi, C. T. (2018). MYC-containing amplicons in acute myeloid leukemia: genomic structures, evolution, and transcriptional consequences. *Leukemia*, *32*(10), 2152–2166. <https://doi.org/10.1038/s41375-018-0033-0>

- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R., Hoad, G., Jang, M., Pakseresht, N., Plaister, S., Radhakrishnan, R., Reddy, K., Sobhany, S., Ten Hoopen, P., Vaughan, R., ... Cochrane, G. (2011). The European Nucleotide Archive. *Nucleic Acids Research*, *39*(suppl_1), D28–D31. <https://doi.org/10.1093/nar/gkq967>
- Leyto-Cruz, F. (2018). Leucemia Mielóide Aguda. *Revista De Hematología*, *19*(1), 24–40. <https://doi.org/10.1016/j>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, N., Cui, Z., Gao, M., Li, S., Song, M., Wang, Y., Tong, L., Bi, Y., Zhang, Z., Wang, S., Zhou, B., & Yin, Z. (2021). Genetic Polymorphisms of PRNCR1 and Lung Cancer Risk in Chinese Northeast Population: A Case-Control Study and Meta-Analysis. *DNA and Cell Biology*, *40*(1), 132–144. <https://doi.org/10.1089/dna.2020.6194>
- Li, Wei, Zhong, C., Jiao, J., Li, P., Cui, B., Ji, C., & Ma, D. (2017). Characterization of hsa_circ_0004277 as a new biomarker for acute myeloid leukemia via circular RNA profile and bioinformatics analysis. *International Journal of Molecular Sciences*, *18*(3). <https://doi.org/10.3390/ijms18030597>
- Li, Wen-hsiung. (1997). *Molecular evolution* (Oxford University Press (ed.); 1st ed.).
- Liang, K. L., Rishi, L., & Keeshan, K. (2013). Tribbles in acute leukemia. *Blood*, *121*(21), 4265–4270. <https://doi.org/10.1182/blood-2012-12-471300>

- Liu, X.-L., Liu, H.-M., Han, N., Li, F.-H., Sun, F., Fan, D.-M., & Xu, Q. (2019). PCAT1 promotes the proliferative and migratory potentials of ovarian cancer via targeting NEK2. *European Review for Medical and Pharmacological Sciences*, 23(19), 8239–8248. https://doi.org/10.26355/eurrev_201910_19133
- Liu, Y., Yi, Y., Wu, W., Wu, K., & Zhang, W. (2019). Bioinformatics prediction and analysis of hub genes and pathways of three types of gynecological cancer. *Oncology Letters*, 18(1), 617–628. <https://doi.org/10.3892/ol.2019.10371>
- Lu, D., Ding, J., Lu, T., Yarla, N. S., Wu, H., & Xu, B. (2018). Cancer Bioinformatics in Cancer Therapy. *Advances in Proteomics and Bioinformatics*, APBI-111, 1–7. <https://doi.org/10.29011/APBI-111.100011>
- Ma, Y.-Q., Liu, Y.-C., Yin, C.-L., Wang, L.-L., Liu, D.-H., Shi, J.-L., & Dou, L.-P. (2021). [Value of SQLE Expression in Prognostic Evaluation of Patients with Acute Myeloid Leukemia]. *Zhongguo shi yan xue ye xue za zhi*, 29(1), 9–16. <https://doi.org/10.19746/j.cnki.issn.1009-2137.2021.01.002>
- Makałowski, W., Gotea, V., Pande, A., & Makałowska, I. (2019). Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In M. Anisimova (Ed.), *Methods in molecular biology (Clifton, N.J.)* (pp. 177–207). Springer New York. https://doi.org/10.1007/978-1-4939-9074-0_6
- Mangul, S., Martin, L. S., Langmead, B., Sanchez-Galan, J. E., Toma, I., Hormozdiari, F., Pevzner, P., & Eskin, E. (2019). How bioinformatics and open data can boost basic science in countries and universities with limited resources. *Nature Biotechnology*, 37(3), 324–326. <https://doi.org/10.1038/s41587-019-0053-y>
- Mashima, J., Kodama, Y., Fujisawa, T., Katayama, T., Okuda, Y., Kaminuma, E., Ogasawara,

- O., Okubo, K., Nakamura, Y., & Takagi, T. (2017). DNA Data Bank of Japan. *Nucleic Acids Research*, *45*(D1), D25–D31. <https://doi.org/10.1093/nar/gkw1001>
- Mohiyuddin, M., Mu, J. C., Li, J., Bani Asadi, N., Gerstein, M. B., Abyzov, A., Wong, W. H., & Lam, H. Y. K. (2015). MetaSV: An accurate and integrative structural-variant caller for next generation sequencing. *Bioinformatics*, *31*(16), 2741–2744. <https://doi.org/10.1093/bioinformatics/btv204>
- Mughal, M. K., Akhter, A., Street, L., Pournazari, P., Shabani-Rad, M.-T., & Mansoor, A. (2017). Acute myeloid leukaemia: expression of MYC protein and its association with cytogenetic risk profile and overall survival. *Hematological Oncology*, *35*(3), 350–356. <https://doi.org/10.1002/hon.2279>
- Nakamura, T. (2015). The role of Trib1 in myeloid leukaemogenesis and differentiation. *Biochemical Society Transactions*, *43*(5), 1104–1107. <https://doi.org/10.1042/BST20150110>
- NIH National Cancer Institute. (2020). Cancer Facts & Figures 2020. *CA: A Cancer Journal for Clinicians*, 1–76.
- Ohanian, M., Rozovski, U., Kanagal-Shamanna, R., Abruzzo, L. V., Loghavi, S., Kadia, T., Futreal, A., Bhalla, K., Zuo, Z., Huh, Y. O., Post, S. M., Ruvolo, P., Garcia-Manero, G., Andreeff, M., Kornblau, S., Borthakur, G., Hu, P., Medeiros, L. J., Takahashi, K., ... Cortes, J. E. (2019). MYC protein expression is an important prognostic factor in acute myeloid leukemia. *Leukemia & Lymphoma*, *60*(1), 37–48. <https://doi.org/10.1080/10428194.2018.1464158>
- Orozco-Arias, S., Isaza, G., & Guyot, R. (2019). Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning.

- International Journal of Molecular Sciences*, 20(15), 3837.
<https://doi.org/10.3390/ijms20153837>
- Orozco-Arias, S., Tobon-Orozco, N., Piña, J. S., Jiménez-Varón, C. F., Tabares-Soto, R., & Guyot, R. (2020). TIP_finder: An HPC Software to Detect Transposable Element Insertion Polymorphisms in Large Genomic Datasets. *Biology*, 9(9), 281.
<https://doi.org/10.3390/biology9090281>
- Pan, Y., Zhan, L., Chen, L., Zhang, H., Sun, C., & Xing, C. (2018). POU5F1B promotes hepatocellular carcinoma proliferation by activating AKT. *Biomedicine & Pharmacotherapy*, 100, 374–380. <https://doi.org/10.1016/j.biopha.2018.02.023>
- Peng, W., & Jiang, A. (2016). Long noncoding RNA CCDC26 as a potential predictor biomarker contributes to tumorigenesis in pancreatic cancer. *Biomedicine & Pharmacotherapy*, 83, 712–717. <https://doi.org/10.1016/j.biopha.2016.06.059>
- Polanski, A., & Kimmel, M. (2007). *Bioinformatics* (Springer Science & Business Media (ed.); 1st ed.). Springer Berlin Heidelberg.
<https://books.google.com.co/books?id=oZbR3GEdmVMC>
- Qu, C.-X., Shi, X.-C., Zai, L.-Q., Bi, H., & Yang, Q. (2019). LncRNA CASC19 promotes the proliferation, migration and invasion of non-small cell lung carcinoma via regulating miRNA-130b-3p. *European Review for Medical and Pharmacological Sciences*, 23(3 Suppl), 247–255. https://doi.org/10.26355/eurrev_201908_18654
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
<https://doi.org/10.1093/bioinformatics/btq033>
- Redaelli, A., Lee, J. M., Stephens, J. M., & Pashos, C. L. (2003). Epidemiology and clinical

- burden of acute myeloid leukemia. *Expert Review of Anticancer Therapy*, 3(5), 695–710.
<https://doi.org/10.1586/14737140.3.5.695>
- Reis, B. S., Jungbluth, A. A., Frosina, D., Holz, M., Ritter, E., Nakayama, E., Ishida, T., Obata, Y., Carver, B., Scher, H., Scardino, P. T., Slovin, S., Subudhi, S. K., Reuter, V. E., Savage, C., Allison, J. P., Melamed, J., Jäger, E., Ritter, G., ... Gnjatic, S. (2013). Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus GAG protein. *Clinical Cancer Research*, 19(22), 6112–6125.
<https://doi.org/10.1158/1078-0432.CCR-12-3580>
- Reyes, M.-H., Pardo, C., & Cuervo, S. (2016). Boletín Epidemiológico INC. *Instituto Nacional de Cancerología - ESE*, 9, 18. <https://www.cancer.gov.co/files/libros/archivos/Boletín No. 9 2016.pdf>
- Rong, L., Huang, W., Tian, S., Chi, X., Zhao, P., & Liu, F. (2018). COL1A2 is a Novel Biomarker to Improve Clinical Prediction in Human Gastric Cancer: Integrating Bioinformatics and Meta-Analysis. *Pathology and Oncology Research*, 24(1), 129–134.
<https://doi.org/10.1007/s12253-017-0223-5>
- Röthlisberger, B., Heizmann, M., Bargetzi, M. J., & Huber, A. R. (2007). TRIB1 overexpression in acute myeloid leukemia. *Cancer Genetics and Cytogenetics*, 176(1), 58–60.
<https://doi.org/10.1016/j.cancergencyto.2007.03.003>
- Sequence Read Archive Submissions Staff. (2011). Using the SRA Toolkit to convert .sra files into other formats. *SRA Knowledge Base [Internet]*, Bethesda(MD).
- Shao, N., Tang, H., Qu, Y., Wan, F., & Ye, D. (2019). Development and validation of lncRNAs-based nomogram for prediction of biochemical recurrence in prostate cancer by bioinformatics analysis. *Journal of Cancer*, 10(13), 2927–2934.

<https://doi.org/10.7150/jca.31132>

Shin, W., Lee, J., Son, S.-Y., Ahn, K., Kim, H.-S., & Han, K. (2013). Human-Specific HERV-K Insertion Causes Genomic Variations in the Human Genome. *PLOS ONE*, 8(4), e60605.

<https://doi.org/10.1371/journal.pone.0060605>

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews. Genetics*, 8(4), 272–285. <https://doi.org/10.1038/nrg2072>

Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L., & Dang, C. V. (2015). MYC, Metabolism, and Cancer. *Cancer Discovery*, 5(10), 1024–1039.

<https://doi.org/10.1158/2159-8290.CD-15-0507>

Stone, R. M., O'Donnell, M. R., & Sekeres, M. A. (2004). Acute Myeloid Leukemia. *Hematology*, 2004(1), 98–117. <https://doi.org/10.1182/asheducation-2004.1.98>

Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology*, 8(1), 90. <https://doi.org/10.1186/1742-4690-8-90>

Tavakolian, S., Goudarzi, H., & Faghihloo, E. (2019). Evaluating the expression level of HERV-K env, np9, rec and gag in breast tissue. *Infectious Agents and Cancer*, 14(1), 1–5.

<https://doi.org/10.1186/s13027-019-0260-7>

Tavakolian, S., Goudarzi, H., Lak, E., & Faghihloo, E. (2020). The evaluation of HERV-K env, np9, rec, gag expression in normal, polyp and cancerous tissues of gastric and colon. *Future Virology*, 14(12), 805–812. <https://doi.org/10.2217/fvl-2019-0114>

Teerlink, C. C., Leongamornlert, D., Dadaev, T., Thomas, A., Farnham, J., Stephenson, R. A., Riska, S., McDonnell, S. K., Schaid, D. J., Catalona, W. J., Zheng, S. L., Cooney, K. A.,

- Ray, A. M., Zuhlke, K. A., Lange, E. M., Giles, G. G., Southey, M. C., Fitzgerald, L. M., Rinckleb, A., ... Cannon-Albright, L. (2016). Genome-wide association of familial prostate cancer cases identifies evidence for a rare segregating haplotype at 8q24.21. *Human Genetics*, 135(8), 923–938. <https://doi.org/10.1007/s00439-016-1690-6>
- Todorovska, E. (2007). Retrotransposons and their Role in Plant—Genome Evolution. *Biotechnology & Biotechnological Equipment*, 21(3), 294–305. <https://doi.org/10.1080/13102818.2007.10817464>
- Tubio, J. (2009). *Evolución de los retrotransposones con LTRs del grupo “Ty3-gypsy” en los genomas de “Anopheles gambiae, Aedes aegypti y Drosophila melanogaster.”* Universidade de Santiago de Compostela.
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, 55(4), 641–658. <https://doi.org/10.1373/clinchem.2008.112789>
- Wallace, T. A., Downey, R. F., Seufert, C. J., Schetter, A., Dorsey, T. H., Johnson, C. A., Goldman, R., Loffredo, C. A., Yan, P., Sullivan, F. J., Giles, F. J., Wang-Johanning, F., Ambs, S., & Glynn, S. A. (2014). Elevated HERV-K mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older men and smokers. *Carcinogenesis*, 35(9), 2074–2083. <https://doi.org/10.1093/carcin/bgu114>
- Wang-Johanning, F., Liu, J., Rycaj, K., Huang, M., Tsai, K., Rosen, D. G., Chen, D. T., Lu, D. W., Barnhart, K. F., & Johanning, G. L. (2007). Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *International Journal of Cancer*, 120(1), 81–90. <https://doi.org/10.1002/ijc.22256>
- Wang, S., Hui, Y., Li, X., & Jia, Q. (2018). Silencing of lncRNA CCDC26 Restrains the Growth

- and Migration of Glioma Cells In Vitro and In Vivo via Targeting miR-203. *Oncology Research*, 26(8), 1143–1154. <https://doi.org/10.3727/096504017X14965095236521>
- Wang, W.-J., Guo, C.-A., Li, R., Xu, Z.-P., Yu, J.-P., Ye, Y., Zhao, J., Wang, J., Wang, W.-A., Zhang, A., Li, H.-T., Wang, C., & Liu, H.-B. (2019). Long non-coding RNA CASC19 is associated with the progression and prognosis of advanced gastric cancer. *Aging*, 11(15), 5829–5847. <https://doi.org/10.18632/aging.102190>
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982. <https://doi.org/10.1038/nrg2165>
- Wong, N., Gu, Y., Kapoor, A., Lin, X., Ojo, D., Wei, F., Yan, J., de Melo, J., Major, P., Wood, G., Aziz, T., Cutz, J.-C., Bonert, M., Patterson, A. J., & Tang, D. (2017). Upregulation of FAM84B during prostate cancer progression. *Oncotarget*, 8(12), 19218–19235. <https://doi.org/10.18632/oncotarget.15168>
- Xue, K., Cao, J., Wang, Y., Zhao, X., Yu, D., Jin, C., & Xu, C. (2020). Identification of Potential Therapeutic Gene Markers in Nasopharyngeal Carcinoma Based on Bioinformatics Analysis. *Clinical and Translational Science*, 13(2), 265–274. <https://doi.org/10.1111/cts.12690>
- Yan, J., Chen, D., Chen, X., Sun, X., Dong, Q., Hu, C., Zhou, F., & Chen, W. (2019). Downregulation of lncRNA CCDC26 contributes to imatinib resistance in human gastrointestinal stromal tumors through IGF-1R upregulation. *Brazilian Journal of Medical and Biological Research*, 52(6), e8399. <https://doi.org/10.1590/1414-431x20198399>
- Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R.,

- Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., ... Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, *48*(D1), D682–D688. <https://doi.org/10.1093/nar/gkz966>
- Yi, J., Zhou, L.-Y., Yi, Y.-Y., Zhu, X., Su, X.-Y., Zhao, Q., Lin, J., Qian, J., & Deng, Z.-Q. (2019). Low Expression of Pseudogene POU5F1B Affects Diagnosis and Prognosis in Acute Myeloid Leukemia (AML). *Medical Science Monitor*, *25*, 4952–4959. <https://doi.org/10.12659/MSM.914352>
- Yokoyama, T., Kanno, Y., Yamazaki, Y., Takahara, T., Miyata, S., & Nakamura, T. (2010). Trib1 links the MEK1/ERK pathway in myeloid leukemogenesis. *Blood*, *116*(15), 2768–2775. <https://doi.org/10.1182/blood-2009-10-246264>
- Yu, Y., Ouyang, Y., & Yao, W. (2018). shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics*, *34*(7), 1229–1231. <https://doi.org/10.1093/bioinformatics/btx763>
- Yuan, Y., Wang, Q., Ma, S. L., Xu, L. Q., Liu, M. Y., Han, B., Du, N., Sun, X. L., Yin, X. L., & Cao, F. F. (2019). lncRNA PCAT-1 interacting with FZD6 contributes to the malignancy of acute myeloid leukemia cells through activating Wnt/ β -catenin signaling pathway. *American Journal of Translational Research*, *11*(11), 7104–7114. <https://pubmed.ncbi.nlm.nih.gov/31814913>
- Yun, S., Sharma, R., Chan, O., Vincelette, N. D., Sallman, D. A., Sweet, K., Padron, E., Komrokji, R., Lancet, J. E., Abraham, I., Moscinski, L. C., Cleveland, J. L., List, A. F., & Zhang, L. (2019). Prognostic significance of MYC oncoprotein expression on survival outcome in patients with acute myeloid leukemia with myelodysplasia related changes (AML-MRC). *Leukemia Research*, *84*, 106194.

<https://doi.org/10.1016/j.leukres.2019.106194>

Zeng, L., Pederson, S. M., Kortschak, R. D., & Adelson, D. L. (2018). Transposable elements and gene expression during the evolution of amniotes. *Mobile DNA*, 9(1), 17. <https://doi.org/10.1186/s13100-018-0124-5>

Zhang, P., Liu, Y., Fu, C., Wang, C., Duan, X., Zou, W., & Zhao, T. (2019). Knockdown of long non-coding RNA PCAT1 in glioma stem cells promotes radiation sensitivity. *Medical Molecular Morphology*, 52(2), 114–122. <https://doi.org/10.1007/s00795-018-0209-8>

Zhang, Y., Li, Q., Yu, S., Zhu, C., Zhang, Z., Cao, H., & Xu, J. (2019). Long non-coding RNA FAM84B-AS promotes resistance of gastric cancer to platinum drugs through inhibition of FAM84B expression. *Biochemical and Biophysical Research Communications*, 509(3), 753–762. <https://doi.org/10.1016/j.bbrc.2018.12.177>

Zhu, X., Zhao, Q., Su, X., Ke, J., Yi, Y., Yi, J., Lin, J., Qian, J., & Deng, Z. (2020). A three-gene signature might predict prognosis in patients with acute myeloid leukemia. *Bioscience Reports*, 40(6), BSR20193808. <https://doi.org/10.1042/BSR20193808>