



Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Biológicas - Programa de Doctorado en Ciencias Biológicas
área Biología Celular y Molecular

LAS REGIONES DE QUIEBRE CROMOSOMAL PRESENTES EN EL LOCUS DEL GEN *RUNX1* PRESENTAN MODULOS CIS REGULADORES ACTIVOS

Tesis para optar al grado de Doctor en Ciencias Biológicas
área Biología Celular y Molecular

MARCELA ALICIA HINOJOSA MORENO
Concepcion – Chile
2018

Profesor Guía: Dra.Soraya Gutiérrez Gallegos
Dpto. Bioquímica y Biología Molecular
Facultad de Ciencias Biológicas
Universidad de Concepción

Esta tesis ha sido realizada en el Departamento de Bioquímica y Biología Molecular de la Facultad de Ciencias Biológicas, Universidad de Concepción.

Profesores integrantes Comisión Evaluadora:

Dra. Soraya Gutiérrez Gallegos
Profesor Guía
Facultad de Ciencias Biológicas

Dra. Violeta Morín Muñoz
Facultad de Ciencias Biológicas



Dr. José Leonardo Gutiérrez C
Facultad de Ciencias Biológicas

Dr. Marcos Alvarez
Profesor Evaluador Externo
Universidad Andres Bello

Dr. Juan Pablo Henríquez H.
Director
Programa Doctorado en Ciencias Biológicas
Área Biología Celular y Molecular

RESUMEN

RUNX1 es un factor de transcripción esencial en el desarrollo de la hematopoyesis y blanco frecuente de translocaciones cromosómicas asociadas al desarrollo de leucemias. A la fecha, el gen *RUNX1* ha sido identificado en alrededor de 65 translocaciones distintas, donde las más frecuentes son las translocaciones cromosómicas 8;21, 12;21 y 3;21. En estas translocaciones, el gen *RUNX1* recombina con los genes *ETO*, *EVI* y *TEL*, respectivamente, en regiones específicas de su locus. Si bien los mecanismos responsables de estos eventos recombinantes son desconocidos, al estudiar los sitios de quiebre mapeados para el gen *RUNX1* se han identificado algunas características comunes: a).- se localizan en regiones acotadas de los intrones, b).- algunas de ellas presentan elementos estructurales de la cromatina como sitios de Hipersensibilidad a Dnasa I, c).- presentan patrones específicos de modificaciones de histonas, d).- Colocalizan o se encuentran próximas a regiones de corte para Topoisomerasa II. Algunas de estas características (b, c y d) han sido asociadas a la presencia de módulos *cis* reguladores.

Con el objetivo de identificar potenciales elementos *cis* reguladores transcripcionalmente activos en regiones de quiebre cromosomal descritas para el gen *RUNX1*, en esta tesis evaluamos el enriquecimiento de distintas marcas epigenéticas asociadas a módulos *cis* reguladores transcripcionalmente activos (H3K27ac, H3K4me1, H3K4me3, H2A.Z), a lo largo del locus del gen *RUNX1*. Nuestros resultados identificaron 4 módulos *cis* reguladores putativos

localizados en las regiones de quiebre cromosomal de los intrones 1 y 5 del gen *RUNX1*. Para validar nuestros resultados se caracterizó mediante ensayos de reportero una de las regiones identificadas (ubicada en el sitio de quiebre del intrón 5 del gen *RUNX1* y denominada región ERP) y evidenciamos que ERP modula la expresión de un gen reportero de luciferasa de manera orientación dependiente. Además, al evaluar la secuencia de ERP encontramos tres sitios de unión a proteínas RUNX. La unión de RUNX1 a estas secuencias se evaluó mediante CHIP, y se encontró que RUNX1 se asocia a la región ERP y forma parte de un complejo enzimático con la histona acetilasa p300. De hecho, al realizar cotransfecciones utilizando un vector de expresión para RUNX1 se evidenció que a altas concentraciones de RUNX1, este aumenta la expresión de un gen reportero con respecto a la actividad basal de la secuencia sin RUNX1.

En conclusión, nuestros resultados demuestran que en una de las regiones de quiebre cromosomal descritas para el gen *RUNX1* se localiza un elemento *cis* regulador que a su vez es controlado por el factor de transcripción RUNX1.

ABSTRACT

RUNX1 is a key regulatory factor for the normal development of the hematopoietic system and frequent target of leukemia associated chromosomal translocations. To date, *RUNX1* gene has been identified in 65 different translocations, being the most frequent t(8;21), t(12;21) and t(3;21) where RUNX1 recombines with ETO, EVI and TEL genes, respectively.

To date, mechanisms responsible of these events are unknown, but studies focusing on Chromosomal Break Regions (BCRs) of *RUNX1* gene have shown some common characteristics: a).- BCRs are harbored at specific intron regions, b).- Some of them exhibit chromatin structural features like DNase I hypersensitive sites and Topoisomerase II cleavage sites c).- Presence of specific patterns of histone modifications. Some of these characteristics (b and c) have been widely associated to the presence of *cis* regulatory modules.

With the purpose of identifying putative *cis* regulatory elements in BCRs described for *RUNX1* gene, in this thesis we evaluated enrichment of the H3K27ac, H3K4me1, H3K4me3 and H2A.Z epigenetics marks throughout the *RUNX1* locus. Our analyses identified 4 putative *cis* regulatory elements mapped in BCRs of introns 1 and 5 of the *RUNX1* gene.

To validate our results, we analyzed one of these regions (at intron 5 defined as ERP) by performing reporter assays, finding that it activates expression of the reporter gene luciferase in an orientation dependent manner. Moreover, when we analyzed putative transcription factors binding sites at ERP, we found three

RUNX binding motifs. To evaluate RUNX1 binding to these motifs we performed Chromatin Immunoprecipitation assays. Our results showed that RUNX1 binds to the ERP region. Indeed when we cotransfected a reporter containing this promoter with a RUNX1 expression vector, we found that RUNX1 modulates the expression of ERP.

Taken together our results demonstrate the presence of a putative regulatory region inside a chromosomal break point cluster in intron 5 of the *RUNX1* gene, which is controlled by RUNX1.



TABLA DE CONTENIDOS

RESUMEN.....	III
ABSTRACT.....	V
INDICE DE FIGURAS.....	IX
INDICE DE TABLAS.....	XII
INTRODUCCION	1
• <i>Organización de la información génica.....</i>	<i>1</i>
• <i>Acetilación y Metilación de Histonas.....</i>	<i>3</i>
• <i>Elementos cis reguladores co-localizan con marcas epigenéticas.....</i>	<i>5</i>
• <i>Hematopoyesis.....</i>	<i>7</i>
• <i>Translocaciones cromosómicas asociadas al desarrollo de leucemias.....</i>	<i>9</i>
• <i>RUNX1: Translocaciones cromosómicas asociadas.....</i>	<i>13</i>
HIPOTESIS.....	19
OBJETIVOS	20
OBJETIVO GENERAL:	20
OBJETIVOS ESPECÍFICOS:	20
MATERIALES Y METODOS.....	21
• <i>Búsqueda de Data Bioinformática.....</i>	<i>21</i>
• <i>Alineamiento con genomas.....</i>	<i>21</i>
• <i>Mapeo de peaks de asociación a lo largo del genoma.....</i>	<i>21</i>
• <i>Caracterización Genómica de peaks.....</i>	<i>22</i>
• <i>Clonamiento.....</i>	<i>22</i>
• <i>Cultivo celular</i>	<i>23</i>
• <i>Transfecciones.....</i>	<i>24</i>
• <i>Inmunoprecipitación de Cromatina (ChIP) y ReChIP.....</i>	<i>24</i>
• <i>Purificación de ADN desde ChIP</i>	<i>26</i>
• <i>PCR en tiempo real.....</i>	<i>27</i>

RESULTADOS.....	28
OBJETIVO ESPECÍFICO N1: “DETERMINAR EL ENRIQUECIMIENTO DE H3K27AC, H3K4ME1, H3K4ME3, H2A.Z A LO LARGO DEL LOCUS DEL GEN <i>RUNX1</i> ”	28
• <i>A nivel genómico, la mayor proporción de las regiones promotoras se encuentran enriquecidas en H3K27ac.....</i>	<i>28</i>
• <i>A nivel genómico, la mayor proporción de las regiones promotoras y exónicas se encuentran enriquecidas en H3K4me3 y H3K4me1.....</i>	<i>36</i>
• <i>A nivel genómico, la mayor parte de las regiones enriquecidas en H2A.Z se localiza en regiones promotoras.....</i>	<i>48</i>
• <i>El locus del gen RUNX1 presenta múltiples regiones enriquecidas en marcas epigenéticas asociadas a módulos cis reguladores.....</i>	<i>53</i>
OBJETIVO ESPECIFICO N2: “EVALUAR LAS MARCAS EPIGENÉTICAS DE H3K27AC, H3K4ME1, H3K4ME3, H2A.Z EN REGIONES DE QUIEBRE CROMOSOMAL.”	59
• <i>Identificación de módulos reguladores cis en regiones de quiebre cromosomal...59</i>	<i>59</i>
OBJETIVO ESPECIFICO N3: “IDENTIFICAR Y CARACTERIZAR POTENCIALES MÓDULOS <i>CIS</i> REGULATORIOS EN EL LOCUS DE <i>RUNX1</i> .”	64
• <i>ERP presenta actividad promotora.....</i>	<i>64</i>
• <i>Identificación in silico e in vivo de un potencial ARN expresado desde el promotor ERP</i>	<i>68</i>
• <i>La proteína RUNX1 se asocia al elemento regulador ERP en regiones de quiebre cromosomal.....</i>	<i>70</i>
• <i>Las HATS pCAF, CBP y p300 se localizan en regiones de quiebre cromosomal y en la región ERP</i>	<i>76</i>
• <i>En la región ERP, RUNX1 y p300 forman parte del mismo complejo.....</i>	<i>77</i>
• <i>RUNX1 modula la actividad transcripcional de la región ERP.....</i>	<i>81</i>
DISCUSION	85
BIBLIOGRAFIA	96

INDICE DE FIGURAS

Figura 1: Representación esquemática del mecanismo de formación de translocaciones	12
Figura 2A: Regiones Intragénicas presentan enriquecimiento en H3K27ac...	31
Figura 2B: Regiones intrónicas presentan un mayor enriquecimiento en H3K27ac.....	32
Figura 3: La mayor proporción de las regiones promotoras, se encuentran efectivamente enriquecidas en H3K27ac.....	35
Figura 4A: A nivel genómico, la mayor parte de las regiones enriquecidas en H3K4me3 se localiza en regiones intragénicas.....	39
Figura 4B: De las regiones intragénicas enriquecidas en H3K4me3, la mayor parte corresponde a regiones exónicas.....	40
Figura 5: La mayor proporción de las regiones promotoras y exónicas, se encuentran efectivamente enriquecidas en H3K4me3.....	42
Figura 6A: A nivel genómico, la mayor parte de las regiones enriquecidas en H3K4me1 se localiza en regiones intergénicas.....	44
Figura 6B: De las regiones intragénicas enriquecidas en H3K4me1, la mayor parte corresponde a regiones exónicas.....	45
Figura 7: La mayor proporción de las regiones promotoras y exónicas, se encuentran efectivamente enriquecidas en H3K4me1.....	47
Figura 8A: A nivel genómico, la mayor parte de las regiones enriquecidas en H2A.Z se localiza en regiones promotoras.....	50

Figura 8B: De las regiones intragénicas enriquecidas en H2A.Z, la mayor parte corresponde a intrones.....	51
Figura 9: La mayor proporción de las regiones promotoras, se encuentran efectivamente enriquecidas en H2A.Z.....	54
Figura 10: El locus del gen <i>RUNX1</i> presenta marcas epigenéticas asociadas a módulos <i>cis</i> reguladores.....	57
Figura 11: El promotor P1 Y P2 del gen <i>RUNX1</i> presenta marcas epigenéticas asociadas a regiones promotoras.....	58
Figura 12: Regiones de quiebre cromosomal presentan enriquecimiento en la acetilación de histona H3 y H4.	62
Figura 13: El intrón 5 del gen <i>RUNX1</i> presenta marcas epigenéticas asociadas a módulos <i>cis</i> reguladores.....	63
Figura 14: Secuencia y características de la región clonada.....	65
Figura 15: La región ERP modula la expresión del gen reportero de luciferasa de manera orientación dependiente.....	67
Figura 16: Identificación <i>in silico</i> de un ARN putativo expresado desde la región ERP.....	69
Figura 17: NAT <i>RUNX1</i> es expresado en células hematopoyéticas.	71
Figura 18: Distribución de sitios putativos de unión a <i>RUNX</i> identificados <i>in silico</i>	74
Figura 19: La proteína <i>RUNX1</i> se asocia a la región ERP en el intrón 5 del gen <i>RUNX1</i>	75

Figura 20: pCAF se asocia a regiones acotadas del intrón 5 del gen *RUNX1*.....**78**

Figura 21: CBP se asocia a regiones acotadas del intrón 5 del gen *RUNX1*.....**79**

Figura 22: p300 se asocia a la región ERP en el intrón 5 del gen *RUNX1*.....**80**

Figura 23: La proteína RUNX1 y p300 forman parte del mismo complejo en el promotor ERP.....**82**

Figura 24: La proteína RUNX1 modula la actividad transcripcional de la región ERP.....**84**



INDICE DE TABLAS

Tabla 1: Translocaciones cromosómicas caracterizadas para el gen <i>RUNX1</i>	17
Tabla 2: Caracterización genómica de las regiones enriquecidas en H3K27ac.....	30
Tabla 3: Normalización genómica del enriquecimiento en H3K27ac.....	34
Tabla 4: Caracterización genómica de las regiones enriquecidas en H3K4me3.....	38
Tabla 5 : Normalización genómica del enriquecimiento en H3K4me3.....	41
Tabla 6: Caracterización genómica de las regiones enriquecidas en H3K4me1.....	43
Tabla 7: Normalización genómica del enriquecimiento en H3K4me1.....	46
Tabla 8: Caracterización genómica de las regiones enriquecidas en H2A.Z.....	49
Tabla 9: Normalización genómica del enriquecimiento en H2A.Z.....	52
Tabla 10: Set de oligonucleótidos intrón 5 del gen <i>RUNX1</i>	61

INTRODUCCION

Organización de la información génica

La expresión génica de cada organismo eucarionte está vinculada estrechamente con el estado de la cromatina, una estructura compleja que no sólo sirve para compactar el genoma en el núcleo, sino que también juega roles importantes en la regulación de la expresión génica *in vivo* (Kornberg, 2007).

La unidad fundamental de compactación de la cromatina es el nucleosoma, cuya conformación estructural fue primero dilucidada por Kornberg (Kornberg 1974) y posteriormente resuelta a nivel atómico por Luger y cols (Luger, Mader et al. 1997). La estructura atómica resultante determinó que un nucleosoma está compuesto por un octámero de proteínas básicas, denominadas histonas, en las que el ADN se enrolla 1,7 veces y organiza 147 pares de bases (pb). Respecto a la conformación del octámero de histonas, esta se caracteriza por un heterotetrámero central de histonas H3-H4 y de dos heterodímeros periféricos de histonas H2A y H2B. Un grado adicional de compactación que puede o no estar presente, se logra mediante la adición de una quinta histona, denominada histona *linker* o histona H1, la cual organiza 20 pb de ADN adicionales y sella finalmente la estructura. Además, entre los nucleosomas es posible encontrar segmentos cortos de ADN que van de 10 a 80 pb, los que se denominan ADN *linker* o ADN de unión (Luger, Mader et al. 1997).

Estructuralmente las histonas presentan dos dominios, un dominio globular que media interacciones histona-histona e histona-ADN y las denominadas “colas de las histonas”, que corresponden en su mayoría a los extremos amino terminal, ricos en aminoácidos básicos, que se extienden desde el *core* nucleosomal hacia el exterior (Luger, Mader et al. 1997, Luger, Dechassa et al. 2012). Si bien las colas de las histonas no contribuyen a la estructura individual de un nucleosoma, de igual modo son fundamentales en el plegamiento internucleosomal (Ausio, Dong et al. 1989, Tse and Hansen 1997, Dorigo, Schalch et al. 2003).

Las colas de las histonas pueden sufrir una serie de modificaciones post traduccionales, entre las que podemos mencionar metilación (Eissenberg and Shilatifard 2010), fosforilación (Andreoli and Del Rio 2014), acetilación (Guillemette, Drogaris et al. 2011), entre otras. Estas modificaciones han sido la base para la formulación del denominado Código de Histonas, que propone que las combinaciones de modificaciones en residuos específicos presentes en la cola de las histonas, son marcadores epigenéticos para la maquinaria enzimática de determinados procesos celulares (Strahl and Allis 2000, Rothbart and Strahl 2014). Por ejemplo, análisis *in vivo* han demostrado que la combinación de acetilación de la lisina 8 de la histona H4 (H4K8) y lisina 14 de la histona H3 (H3K14) se encuentran presentes en regiones transcripcionalmente activas, mientras que la trimetilación de la lisina 9 y 27 de la histona H3 se localizan en regiones transcripcionalmente reprimidas (Alvarez-Venegas and Avramova 2005, Eissenberg and Shilatifard 2010, Guillemette, Drogaris et al.

2011). Recientemente, y gracias al avance de técnicas de secuenciación de genomas completos se ha establecido que una importante marca epigenética corresponde a la H3K27ac que se encuentra presente en regiones transcripcionalmente activas (Wang, Zang et al. 2016). En contraste, la H3K27me3 se localiza frecuentemente en regiones transcripcionalmente reprimidas y la H3K36me3 ha sido mapeada en el cuerpo del locus de genes que están siendo transcritos (Jiao and Liu 2015, Hosogane, Funayama et al. 2016, Mahrez, Arellano et al. 2016, Li and Wang 2017).

Además del rol de estas modificaciones como señal epigenética, se ha observado que las modificaciones en la cola de las histonas afectan la estructura de orden superior de la cromatina, alterando el contacto entre nucleosomas y las interacciones histona-ADN (Garcia-Ramirez, Dong et al. 1992, Vettese-Dadey, Grant et al. 1996, du Preez and Patterson 2013).

Acetilación y Metilación de Histonas

La acetilación de histonas ha sido ampliamente estudiada por su asociación con la transcripción, debido a que esta disminuye la interacción histona-histona e histona-ADN, lo cual favorece la accesibilidad de la cromatina (Sterner and Berger, 2000; Imhof and Wolffe, 1998). Regiones transcripcionalmente activas se caracterizan por presentar una estructura de cromatina abierta y accesible a la maquinaria transcripcional. De hecho, es frecuente encontrar hiperacetilación de residuos de lisina en estas .

Por ejemplo, análisis de secuenciación masiva han determinado que los promotores de los genes que están siendo transcritos presentan un alto

enriquecimiento en trimetilación de lisina 4 de la histona H3 (H3K4me3) (Barski, Cuddapah et al. 2007, Kimura 2013, Huang, Jiang et al. 2014) y también en la acetilación de la lisina 27 de la Histona H3 (H3K27ac) (Barski, Cuddapah et al. 2007, Heintzman, Stuart et al. 2007). Estudios más recientes han evidenciado que la acetilación de los residuos de lisina 64 y lisina 122 de la histona H3 se localizan preferentemente en promotores y enhancers activos (Pradeepa, Grimes et al. 2016).

Las enzimas encargadas de acetilar las histonas son denominadas Histonas Acetilasas (HATs) y catalizan la transferencia de un grupo de acetilo desde acetilCoA al extremo e-amino de los residuos de lisina (Marmorstein and Trievel, 2009). Estas enzimas poseen una amplia variedad de sustratos (Sternier and Berger, 2000) y además de su función catalítica, estas enzimas son capaces de interactuar con factores basales de la transcripción, tales como TBP, TFIIB y ARN polimerasa II (Sternier and Berger, 2000; Kalkhoven, 2004; Brown et al., 2000), facilitando el proceso de transcripción. En la actualidad varias HATs han sido caracterizadas, de éstas, las más estudiadas son p300, CBP y pCAF, las que han sido identificadas regulando la expresión de genes claves del desarrollo (Schiltz et al., 1999; Ramos et al., 2010; Ogryzko et al., 1996).

CBP y p300 son coactivadores transcripcionales con un alto grado de homología entre sí (Ogryzko et al., 1996; Liu et al., 2008). Generalmente forman un complejo que se asocia a regiones promotoras regulando la expresión génica por acetilación de histonas y factores de transcripción (Liu et al., 2008; Blobel, 2000; Blobel, 2002). Además se caracterizan por actuar como proteína puente

estabilizadora de complejos transcripcionales (Blobel, 2000). Varios estudios han demostrado que ambas se unen a una variedad de factores específicos de transcripción, tales como RUNX, PU.1, GATA, etc. facilitando la expresión de genes linaje específico (Kitabayashi et al., 1998a; Katsumoto et al., 2008; Kalkhoven, 2004; Giordano and Avantaggiati, 1999; Blobel, 2000; Blobel, 2002).

Tanto CBP como p300 también interactúan con pCAF (p300/CBP associated factor), que al igual que CBP y p300 posee actividad HAT intrínseca. PCAF es una proteína de 832 aminoácidos con un dominio carboxiterminal homólogo a la histona acetilasa GCN5 de levadura (Schiltz et al., 1999; Sterner and Berger, 2000)

Elementos cis reguladores co-localizan con marcas epigenéticas

La transcripción es el proceso mediante el cual el ADN es transcrito a ARN, elemento mediador de la expresión génica. Para que este proceso sea llevado a cabo de forma exitosa, es necesaria la acción coordinada de una serie de elementos *cis* y *trans*.

Los elementos *cis* o también conocidos como Módulos Reguladores *Cis* (MRC) (Barrera and Ren 2006, Heinz and Glass 2012, Beck, Thoms et al. 2013) son secuencias de ADN, generalmente no codificante, presentan sitios de unión a factores de transcripción y/u otras moléculas regulatorias necesarias para la transcripción. Promotores y enhancers son dos de los MRCs más estudiados (Rye, Saetrom et al. 2011, Calero-Nieto, Ng et al. 2014). Los primeros actúan de manera orientación dependiente (Grunberg and Hahn 2013, Roy and Singer 2015) y se localizan inmediatamente corriente arriba del sitio de inicio de la

transcripción. Estas regiones permiten el posicionamiento de la ARN polimerasa para iniciar la transcripción (Grunberg and Hahn 2013, Roy and Singer 2015).

Diversos grupos de investigación han determinado que generalmente los promotores y enhancers, se caracterizan por presentar una serie de marcas epigenéticas, elementos estructurales de la cromatina y asociación con factores de transcripción específicos, que en la actualidad permiten su reconocimiento. De hecho, estudios realizados por Barski y cols (Barski, Cuddapah et al. 2007) y por Heintzman y cols (Heintzman, Stuart et al. 2007) identificaron las marcas epigenéticas asociadas a promotores en escala genómica y demostraron que estos se caracterizan por enriquecimiento en una serie de marcas tales como trimetilación de la lisina 4 de la histona H3 (H3K4me3), monometilación de la lisina 4 de la histona H3 (H3K4me) y enriquecimiento en la variante de histona H2A.Z (Alarcon-Riquelme 2003, Barski, Cuddapah et al. 2007, Rangasamy 2013). Adicionalmente, en otro estudio se estableció que en particular, las regiones promotoras presentan acetilación preferente de H3K9ac, H3K23ac y H3K14ac y presencia de la histona acetilasa pCAF (Rajagopal, Ernst et al. 2014).

Por otra parte los enhancers son los responsables de amplificar la expresión génica, actúan independientes de posición y orientación por lo que pueden localizarse cercanos o a gran distancia del gen que modulan (Symmons and Spitz 2013, Heinz, Romanoski et al. 2015). Se caracterizan por presentar marcas epigenéticas muy parecidas a regiones promotoras, como por ejemplo

H3K4me3 y H3K4me y particularmente presentan la diacetilación de las lisinas 9 y 14 de la histona H3 (Barski, Cuddapah et al. 2007, Chen, Morris et al. 2012).

Ambos elementos, promotores y enhancers, regulan la expresión génica dependiendo de la unión *in vivo* de factores de transcripción. Esta unión está determinada por el mayor grado de descompactación de la cromatina, observable por la presencia de elementos estructurales de la cromatina tales como sitios de hipersensibilidad a DNasa I y sitios de corte a Topoisomerasa II. De hecho, varios grupos ya han determinado en distintos organismos y modelos celulares una fuerte asociación entre sitios de hipersensibilidad a DNasa I, patrones de acetilación y metilación de histonas y la unión específica de factores de transcripción con la actividad transcripcional del MRC (Boyle, Davis et al. 2008, Thurman, Rynes et al. 2012).

Hematopoyesis

La hematopoyesis es el mecanismo fisiológico responsable de la formación continuada de los distintos linajes celulares sanguíneos. Se caracteriza por ser un proceso altamente controlado, donde una célula troncal (Hematopoietic Stem Cell o HSC) prolifera, diferencia, madura y libera a la sangre los diferentes tipos celulares sanguíneos. En vertebrados, es posible distinguir distintas fases hematopoyéticas basadas principalmente en el lugar anatómico donde ocurren (Baron and Fraser 2005).

La primera fase, denominada hematopoyesis primitiva se inicia en estadios tempranos del desarrollo humano (E15-18) en una estructura denominada endotelio hemogénico, formada a partir del mesodermo. En esta

fase se generan las primeras células eritroides que contienen cadenas primitivas de globinas y presentan núcleo (Smith and Glomski 1982, Adamo and Garcia-Cardena 2012). La segunda etapa o hematopoyesis fetal se desarrolla en el hígado fetal (E35-42 en humanos) y se caracteriza por la aparición de eritrocitos anucleados con cadenas fetales de globina (Baron and Fraser 2005). Por último, la hematopoyesis definitiva se desarrolla a partir del séptimo mes fetal, se caracteriza porque el principal órgano hematopoyético es la médula ósea (único órgano hematopoyético después del nacimiento) y por la producción masiva de eritrocitos anucleados con cadenas de globina alfa y beta (Baron, Isern et al. 2012).

La coordinación y regulación de las etapas involucradas en el desarrollo normal de la hematopoyesis es compleja. En ella intervienen una serie de factores que regulan la capacidad de proliferar, diferenciar y madurar de una célula hematopoyética. Alteraciones en los elementos de control de estos procesos se han asociado al desarrollo de diversas patologías, siendo una de las más frecuentes la leucemia (DeVilbiss, Sanalkumar et al. 2014).

La leucemia es una enfermedad maligna y clonal del sistema hematopoyético producida por la adquisición y proliferación de progenitores hematopoyéticos que presentan características oncogénicas. Según su evolución, estas pueden ser clasificadas como crónicas o agudas; o bien de acuerdo a la estirpe celular involucrada en mieloides o linfoides (Arber, Orazi et al. 2016).

Se ha sugerido que la exposición a ciertos agentes tóxicos; como bencenos, agentes alquilantes, etc; se asocian al desarrollo de la enfermedad. No obstante, a la fecha no han sido identificadas las causas etiológicas responsables de su desarrollo. Sin embargo, se ha observado que las células oncogénicas presentes en distintos tipos de cáncer, presentan alteraciones genéticas específicas (Rowley 1973, Mitelman 2000) entre las que podemos mencionar deleciones, inserciones y translocaciones cromosómicas, siendo esta última la más frecuente encontrada tanto en tumores líquidos (leucemia) como en tumores sólidos (Vardiman, Harris et al. 2002, Vardiman, Thiele et al. 2009)

De acuerdo a la Clasificación de Neoplasias Hematológicas formulada por la Organización Mundial de la Salud (OMS) a partir del año 2001, se han utilizado alteraciones genéticas para el diagnóstico de pacientes con leucemia, siendo las translocaciones cromosómicas una de las más utilizadas. Así, en los seres humanos el 65% de los casos de leucemia mieloide aguda (LMA) presentan translocaciones cromosómicas (Mitelman 2000, Arber, Orazi et al. 2016). De hecho, la detección de los genes involucrados en la formación de translocaciones cromosómicas ha permitido identificar varios de los genes maestros del sistema hematopoyético, entre los que podemos mencionar *RUNX1*, *RAR*, *EVI*, *FLT3*, *cMyc*, entre otros (Arber, Orazi et al. 2016).

Translocaciones cromosómicas asociadas al desarrollo de leucemias

Las translocaciones cromosómicas generan intercambio de material genético, teóricamente en cualquier loci, entre los brazos de dos cromosomas vecinos. Las translocaciones que se presentan en humanos son de gran interés

clínico debido a que están ligadas de forma específica a un gran número de desórdenes como por ejemplo retardo mental, infertilidad, cáncer y leucemias (Mitelman 2000). De hecho, su estudio ha permitido establecer distintos criterios diagnósticos y de sobrevida para pacientes con cáncer. La OMS el año 2002 (Vardiman, Harris et al. 2002), basándose en la presencia de estas anomalías estableció una serie de protocolos diagnósticos y terapéuticos, que determinan el manejo de distintos tipos de neoplasias. A lo largo de los años, y gracias a esta clasificación, se han establecido distintos indicadores de pronóstico y sobrevivencia a la enfermedad, además de mejoras en las indicaciones quimioterapéuticas que, a la larga, han aumentado los años de sobrevida a la enfermedad (Vardiman, Harris et al. 2002, Vardiman, Thiele et al. 2009, Arber, Orazi et al. 2016).

Si bien diversos estudios han demostrado que las neoplasias son mediadas por translocaciones cromosómicas, a pesar de los esfuerzos realizados los mecanismos moleculares involucrados en la generación de estas aún son desconocidos. No obstante, se ha sugerido que para que se produzca una translocación cromosómica es necesario que ocurra al menos un quiebre en la hebra de ADN en diferentes cromosomas, la yuxtaposición de los cromosomas involucrados y un mecanismo de reparación anómalo. De hecho, se ha determinado que cuando dos cromosomas sufren quiebres de forma simultánea, se aumenta la probabilidad de generar una translocación (Richardson and Jasin 2000). Dependiendo del contenido neto total de ADN posterior a la translocación, estas se clasifican en translocaciones cromosómicas

de tipo recíproco y no recíproco. Una translocación recíproca ocurre cuando los extremos del ADN de cada quiebre son ligados a los extremos del quiebre presente en otro cromosoma como resultado de un mecanismo erróneo de reparación, no involucrando una pérdida o ganancia neta total en el contenido de ADN cromosomal. En contraste, una translocación de tipo no recíproco involucra el intercambio unilateral de material genético entre un cromosoma “dador” y un cromosoma “aceptor” como producto final de esta translocación se produce una pérdida en el contenido de ADN total (Rabkin and Janz 2008) (**Figura 1**).

El estudio de los sitios de quiebre del ADN identificados hasta la fecha, ha permitido evidenciar características importantes, por ejemplo que se localizan en regiones específicas de los genes recombinantes, preferentemente en intrones, y presentan variaciones de tamaño entre sí. Si bien el clon leucémico de cada paciente aparentemente presenta un sitio único de quiebre, no existe un único sitio de recombinación. Este es diferente en distintos pacientes aun cuando todos tienden a estar agrupados a una región relativamente definida a la que se le ha denominado Región de Agrupamiento de Quiebres o BCRs por su sigla en inglés por *Break Cluster Region*. Los BCRs se localizan preferentemente en intrones. Todas estas características nos permiten sugerir por una parte que los BCRs no ocurren de forma azarosa en el ADN y que las regiones de quiebre entre los genes que translocan pueden presentar características comunes que expliquen la especificidad aparente del mecanismo de recombinación involucrado. La búsqueda de estas características ha sido un campo de estudio y debate en los últimos 20 años.

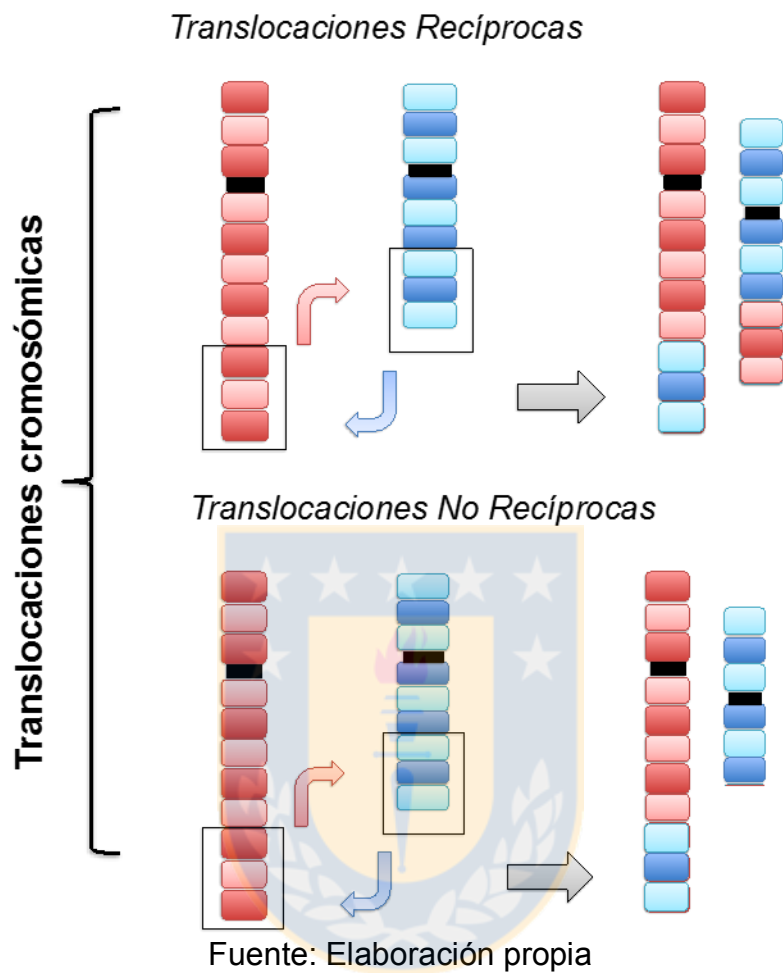


Figura 1: Representación esquemática del mecanismo de formación de translocaciones.

En el panel superior se presenta el mecanismo de formación de translocaciones de tipo recíprocas, donde dos cromosomas (representados de color rojo y azul) intercambian material genético (resaltado en un cuadro negro), originando dos cromosomas quiméricos derivados

En el panel inferior se presenta el mecanismo de formación de translocaciones de tipo no recíprocas, donde al igual que en el caso anterior, se presentan dos cromosomas (azul y rojo) para los cuales existe intercambio de material genético, sin embargo solo uno de ellos acepta esta recombinación, generándose un cromosoma híbrido y un cromosoma de menor tamaño (producto de la pérdida de material genético).

Como primer acercamiento, y tomando en cuenta los mecanismos de recombinación involucrados, se estudió si en los BCRs existían secuencias de ADN comunes, que pudieran sugerir un mecanismo de reparación aberrante. Sin embargo, los resultados demuestran que, en la mayor parte de los casos, no existen secuencias de ADN comunes que permitan explicar este patrón particular de recombinación. No obstante, distintos grupos de investigación han observado la presencia de elementos estructurales de la cromatina presentes en los puntos de quiebre (Zhang, Strissel et al. 2002).

RUNX1: Translocaciones cromosómicas asociadas

Uno de los principales genes involucrados en translocaciones cromosómicas y presente en leucemia es el gen *RUNX1* (Nucifora and Rowley 1995, Arber, Orazi et al. 2016), el cual pertenece a una familia de genes que codifican para factores de transcripción esenciales a lo largo del desarrollo (Cohen, 2009; Ito, 2004; Wang et al., 2009). Ratones *knockout* para este gen, son letales en estadios tempranos de la embriogénesis (E12.5-13.5), debido al bloqueo en el desarrollo del sistema hematopoyético, lo cual ha demostrado que tiene un rol clave en la hematopoyesis (Okuda et al., 1996). Su expresión espacio temporal es regulada a través de dos regiones promotoras, las que generan una amplia variedad de isoformas involucradas en diferentes procesos celulares (Ghozi et al., 1996). Estructuralmente, la proteína *RUNX1* se compone de dos grandes motivos (Cohen, 2009); un dominio de unión al ADN, localizado en la porción aminoterminal y un dominio de transactivación ubicado hacia el extremo carboxilo terminal. El dominio de unión al ADN está

compuesto por 128 aminoácidos altamente conservados a lo largo de la evolución que es homólogo al dominio runt de *Drosophila*, por lo que se le ha denominado *Runt homology domain* (RHD) (Tahirov et al., 2001). Mediante este mismo dominio, RUNX1 interacciona con su subunidad acompañante, la subunidad beta, lo que produce un aumento de su afinidad por el ADN, este heterodímero es el factor de transcripción funcionalmente activo (Perry et al., 2002; Tahirov et al., 2001; Wang et al., 2010; Bartfeld et al., 2002). Por otra parte, el Dominio de Transactivación (TD), interacciona con coactivadores y corepresores transcripcionales (Kitabayashi et al., 2001; Mikhail et al., 2006; Otto et al., 2003; Perry et al., 2002; Wang et al., 2009). De este modo, RUNX1 regula la expresión de genes importantes en hematopoyesis, tales como el gen de mieloperoxidasa (MPO), elastasa, receptor del factor estimulador de colonias de granulocitos y macrófagos, receptor del factor estimulador de colonias de macrófagos y receptor β de células T (Lichtinger et al., 2010; Mikhail et al., 2006, Otto et al., 2003; Oakford et al., 2010; Woolf et al., 2003).

La regulación de la expresión génica mediada por la proteína RUNX1 es compleja, ya que por una parte depende de su unión a secuencias consenso localizadas en regiones reguladoras y, por otra , de su posterior interacción con distintos complejos macromoleculares (Otto et al., 2003; Niebuhr et al., 2008; Miller et al., 2002; Lichtinger et al., 2010). Diversos estudios han demostrado que las proteínas RUNX, para regular la expresión génica deben interactuar con complejos multiproteicos, lo que parece ser crucial para la interacción efectiva

con los sitios de unión en la respectiva región *cis* reguladora (Otto et al., 2003; Mikhail et al., 2006; Lichtinger et al., 2010; Downing, 2003).

Además de lo señalado anteriormente, existe un nivel adicional de complejidad en la regulación de la expresión génica mediada por RUNX1. Según sea el contexto celular RUNX1 puede activar o silenciar la expresión génica (Woolf et al., 2003; Otto et al., 2003; Oakford et al., 2010; Mikhail et al., 2006). Esta dualidad funcional es determinada por su capacidad de reclutar complejos coactivadores o corepresores a las regiones reguladoras *cis* de sus genes blanco (Woolf et al., 2003; Otto et al., 2003; Oakford et al., 2010; Mikhail et al., 2006). De este modo, RUNX1 puede reclutar una serie de enzimas tales como, HATs, histonas deacetilasas (HDAC), factores remodeladores de la cromatina ATP dependientes, los cuales favorecerán la formación de una estructura de cromatina permisiva o inaccesible a la maquinaria transcripcional.

El gen que codifica para la proteína RUNX1 se ubica en el brazo largo del cromosoma 21 y posee un tamaño de 260 kb, distribuidos a lo largo de 9 exones y dos promotores localizados a gran distancia entre ellos que permiten la expresión de distintas isoformas (Levanon et al., 2001). Ahora, desde el punto de vista hematólogico, uno de los genes blanco de translocaciones cromosomales corresponde al gen *RUNX1*, el cual se ha visto involucrado en alrededor de 65 translocaciones distintas. No obstante, a la fecha solo 35 de estos eventos han sido estudiados (Mitelman F 2017) y solo en algunos casos se ha identificado el compañero de recombinación y sitio de quiebre de esta. Es importante destacar que, de las translocaciones estudiadas que afectan al

cromosoma 21 todas recombinan en regiones intrónicas del gen *RUNX1* (**Tabla 1**). Una de las translocaciones que involucra al gen *RUNX1* y de mayor relevancia y frecuencia clínica es la que involucra a los genes *RUNX1* y *ETO*, generando la proteína de fusión quimérica *RUNX1/ETO* (*AML1/ETO*) (Growney et al., 2004; Kurokawa and Hirai, 2003; Okuda et al., 1996; Perry et al., 2002; Xiao et al., 2001). Por esta razón diversos grupos han estudiado en detalle los sitios de quiebre de ambos genes a fin de determinar el mecanismo por el cual ocurre esta translocación (Stuardo et al., 2009; Xiao et al., 2001; Zhang et al., 2002).

Zhang y cols (2002) al mapear los sitios de quiebre en 31 pacientes con leucemias mieloides, encontró que estos se localizan en tres BCR localizados en el intrón 5 del gen *RUNX1* (Zhang et al., 2002). Otros estudios también han demostrado resultados similares, Xiao y cols (Xiao et al., 2001), al estudiar 10 pacientes que presentaron la translocación en leucemias de tipo mieloide, encontraron que los BCRs podían ser agrupados en dos regiones, las cuales coincidían con las descritas por Zhang. Además, se ha evidenciado la presencia de elementos estructurales de cromatina en las regiones cercanas a sitios de quiebre cromosomal, como por ejemplo sitios de hipersensibilidad a Topoisomerasa II y sitios de corte a DNAsa I (Zhang et al., 2002). Estudios previos de nuestro laboratorio publicados por Stuardo y cols (Stuardo, Martinez et al. 2009), también han evidenciado un patrón específico de acetilación caracterizado principalmente por enriquecimiento en la acetilación de histona H3 y H4 en el interior o cercanas a las regiones de quiebre cromosomal.

Tabla 1: Translocaciones cromosómicas caracterizadas para el gen *RUNX1*.

Translocación	Genes Recombinantes		Sitio del quiebre en <i>RUNX1</i>	Fenotipo asociado	Frecuencia según OMS	Referencia
t(8;21)	RUNX1	ETO	Intrón 5	M2	La más recurrente	(Peeters, Buys et al. 1992, Arber, Slovak et al. 2002, Gustafson, Lin et al. 2009, Bae, Kim et al. 2013, Krauth, Eder et al. 2014)
t(3;21)	RUNX1	MECOM	intrón 6	LMA	Recurrente	(Nucifora and Rowley 1995)
t(12;21)	RUNX1	ETV6	Intrón 1	LLA	Recurrente	(Golub, Barker et al. 1995, Wiemels and Greaves 1999, Andreasson, Hoglund et al. 2000, Attarbaschi, Mann et al. 2004, Al-Sweedan, Neglia et al. 2007)
t(1;21)	RUNX1	CLCA2	intrón 6	M2		(Nadal, Stephan et al. 2008, Giguere and Hebert 2010)
t(3;21)	RUNX1	RPL22P1	intrón 6	LMA	Recurrente	(Nucifora and Rowley 1995)
t(7;21)	RUNX1	USP42	intrón 6	MDS - AML		(Foster, Paulsson et al. 2010, Giguere and Hebert 2011)
t(15;21)	RUNX1	TCF12	intrón 6	MDS - AML		PMC4681058
t(11;21)	RUNX1	KIAA1549L	Intrón 5 y 6	LMA		(Abe, Katsumi et al. 2012)
t(2;21)	RUNX1	LAF4	Intrón 6	LLA		(Chinen, Taki et al. 2008)
t(15;21)	RUNX1	SV2B	Intrón 1	LLA		(Giguere and Hebert 2013)
t(21;21)	RUNX1	DSCAM	No determinado	LLA-B		(Gu, Churchman et al. 2016)
t(21;21)	RUNX1	DYRK1A	No determinado	LLA-B		(Gu, Churchman et al. 2016)
t(17;21)	RUNX1	FAM117A	No determinado	LLA-B		(Lilljebjorn, Henningsson et al. 2016)

Fuente: Elaboración propia

Todos estos antecedentes nos sugieren por una parte que la estructura de cromatina de las regiones de quiebre que afectan al gen *RUNX1* son accesibles y presentan características similares a las observadas en regiones *cis* regulatorias, lo que nos permite plantearnos como pregunta biológica ¿Se encontrarán módulos *cis* reguladores en las regiones de quiebre cromosomal presentes en el gen *RUNX1*?



HIPOTESIS

RUNX1 es un factor de transcripción esencial para el desarrollo de la hematopoyesis y es blanco frecuente de translocaciones cromosómicas asociadas al desarrollo de leucemias. El gen *RUNX1* presenta múltiples regiones de quiebre cromosomal localizadas en regiones intrónicas acotadas, para algunas de estas regiones han sido mapeados elementos estructurales de la cromatina, tales como sitios de hipersensibilidad a DNasa I y de corte a topoisomerasa II, y marcas epigenéticas caracterizadas principalmente por hiperacetilación de histona H3 y H4. Estas características se localizan frecuentemente en módulos *cis* reguladores activos, por este motivo planteamos como hipótesis que ***“Las regiones de quiebre cromosomal localizadas en el locus del gen RUNX1 presentan módulos cis reguladores transcripcionalmente activos”***

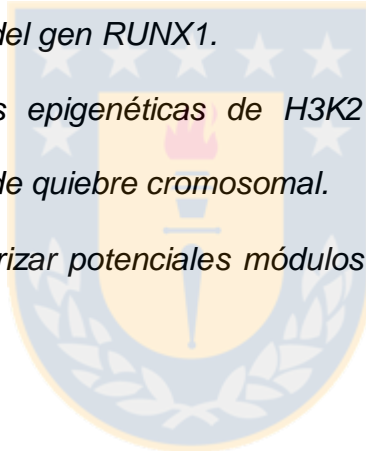
OBJETIVOS

Objetivo General:

- Determinar si las regiones de quiebre cromosomal presentes en el gen RUNX1 presentan módulos *cis* regulatorios transcripcionalmente activos.

Objetivos Específicos:

- *Determinar el enriquecimiento de H3K27ac, H3K4me1, H3K4me3, H2A.Z a lo largo del locus del gen RUNX1.*
- *Evaluar las marcas epigenéticas de H3K27ac, H3K4me1, H3K4me3, H2A.Z en regiones de quiebre cromosomal.*
- *Identificar y caracterizar potenciales módulos *cis* regulatorios en el locus de RUNX1.*



MATERIALES Y METODOS

Búsqueda de Data Bioinformática

Los datos analizados fueron generados mediante ensayos de Inmunoprecipitación de Cromatina asociada a secuenciación masiva (ChIPseq) disponibles en los servidores ENCODE (<http://genome.ucsc.edu>). Se utilizó información generada por diversos grupos para la línea celular linfoblástica GM12878. Las librerías analizadas se localizan en el grupo “Regulation” y los subgrupos ENCODE Regulation, BroadChromHMM, BroadHistone, Open Chromatin entre otros. Los datos analizados corresponden a información accesible universalmente.

Alineamiento con genomas

Posterior a su descarga, la data bruta de ChIPseq (RAW data) fue primeramente alineada contra el genoma humano de referencia 19 (Hg19) utilizando un script escrito en lenguaje Shell. Para ello, se utilizó un valor de ventana (window value=ww) de 44 nucleótidos, correspondientes al tamaño promedio de los fragmentos secuenciados, permitiendo un mismatch de a lo más 2 nucleótidos por secuencia.

Mapeo de peaks de asociación a lo largo del genoma

Para todos los set de datos se utilizó el programa MACS. Para ello se empleó la línea de comando por defecto y se usaron los siguientes parámetros:

mfold = 16, t size = 25 y band weight = 400, pvalue = 1 e-9 y genomic size = 2.700.000.000, que corresponde al resultado de la resta entre el tamaño del genoma humano secuenciado (Hg19: 3200Mb) menos las secuencias repetidas y regiones gaps que no han sido resueltas. La sobreposición fue establecida utilizando 20 nucleótidos como rango de referencia.

Caracterización Genómica de peaks.

Para el análisis de distribución de peaks se utilizó el sistema de herramientas del servidor GALAXY Project (<http://galaxyproject.org>).

Clonamiento

Se utilizó el método de digestión/ligación simultánea. Para ello se amplificó el fragmento de ADN de interés mediante PCR utilizando la enzima *Pfx platinum*, caracterizada por su alta fidelidad en la polimerización de nucleótidos y por generar extremos romo en los productos amplificados. Se diseñó un partidor sentido (5' TCG CTG TCC TGT TCA TT '3) y un partidor antisentido (5' AGG GAG ACA GCC CTC ATA AAC T'3), el producto de amplificación corresponde a un fragmento de 2.1kb. El programa de PCR utilizado fue: denaturación inicial 98°C- 5 minutos, denaturación 98°C- 30 segundos, alineamiento 60°C- 30 segundos, extensión 68°C- 90 segundos, 30 Ciclos; extensión final de 68°C- 5 minutos. Posteriormente el fragmento de PCR fue purificado utilizando el kit PCR cleanup (Macherey Nagel) siguiendo las instrucciones del fabricante. Como vector de clonamiento, se utilizó el plásmido pGL3-basic (Promega). Para el clonamiento, en un volumen final de 20 µl se

adicionó 50 ng de plásmido, 5 unidades enzimáticas de la enzima de restricción SmaI (New England Biolabs), 10 unidades de T4 DNA Ligasa (New England Biolabs), 1X buffer KGB (0.1M Acetato de Potasio, 0.025M Tris pH 7.6, 0.01M Acetato de Magnesio Tetrahidratado, 0.5mM b-mercaptoetanol, 0,01 mg/ml Suero de Albumina Bovina), 2 mM de ATP y 500 ng de inserto. Esta reacción fue incubada durante 2 horas a 25°C, posteriormente 20 minutos a 65°C, seguido de 5 minutos en hielo. Luego, se adicionó 5 unidades de SmaI y se incubó 2 horas a 22°C. Luego, 10 µl de la mezcla de ligación fue adicionada a 100 µl de bacterias competentes DH5alfa, las cuales fueron transformadas mediante shock térmico y sembradas en agar LB 100 µg/ml ampicilina durante 12 horas. Los clones positivos (que incorporaron el inserto) fueron seleccionados mediante Colony-PCR, para lo cual colonias aisladas fueron incubadas en 50 µl de medio LB ampicilina (50 µg/ml) durante 30 minutos. Posteriormente, 2 µl de este cultivo fue utilizado como templado para una reacción de PCR utilizando la enzima Taq DNA polymerase (Invitrogen) utilizando partidores que fueron diseñados rodeando la región de clonamiento múltiple (multiple cloning site = MCS). El programa de PCR utilizado fue: denaturación inicial 95°C por 5 minutos, 35 ciclos de PCR con denaturación 95°C por 30 segundos, alineamiento 60°C por 30 segundos y extensión 72°C por 2 minutos y una extensión final de 5 minutos a 72°C.

Cultivo celular

Se utilizaron las líneas celulares HeLa y HL60 (aisladas desde cáncer cérvico uterino y leucemia promielocítica respectivamente), las que fueron

cultivadas en medio DMEM y RPMI 1640 respectivamente y suplementadas con suero fetal bovino 10% y 0.2 mM de glutamina y los antibióticos Penicilina/Estreptomicina 100 U/ml. Bajo condiciones controladas de temperatura 37°C y 5% CO₂.

Transfecciones

Todas las transfecciones fueron realizadas en células HeLa. Las células fueron transfectadas a un 80% de confluencia en placas de 12 pocillos. Las transfecciones fueron realizadas con el kit NanoJuice (Merck) siguiendo las instrucciones entregadas por el fabricante.

Inmunoprecipitación de Cromatina (ChIP) y ReChIP.

Se empleó el método descrito por E. Soutoglou y Talianidis (Science, 2002) con algunas modificaciones: las células fueron tratadas con formaldehído 1% final por 10 minutos, la reacción de entrecruzamiento se detuvo adicionando glicina a una concentración final de 0.125 M por 5 minutos. Luego las células se lavaron con tampón fosfato salino (PBS) (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄·7H₂O, 1.4 mM KH₂PO₄) frío y se centrifugaron a 4°C por 5 minutos a 165xg. Posteriormente el pellet celular se resuspendió en buffer de lisis (25 mM de Hepes pH 7.8, 1.5 mM MgCl₂, 10 mM de KCl, 0.1% de NP- 40, Cϕmplete 1X (®Roche) y el inhibidor de deacetilasas Butirato 10 mM) y se incubó en hielo por 10 minutos. A continuación, se procedió a romper las células en homogenizador Douncer, 20 golpes con émbolo A. Los núcleos fueron recogidos por centrifugación a 165 x g por 5 minutos a 4°C y resuspendidos en buffer de

sonicación (50 mM de Hepes pH 7.9, 140 mM de NaCl, 1 mM de EDTA, 1% v/v Tritón X-100, 0.1% v/v SDS, 0.1% v/v NP-40), Cϕmplete 1X (®Roche) y butirato (10 mM) y se sonicaron en hielo usando sonicador 3000 de Misonix con Microtip, hasta que se obtuvieron fragmentos de ADN menores a 600 pb. Posteriormente los sonicados se centrifugaron a 16000 x g por 15 minutos a 4°C, se midió la absorbancia del sobrenadante a 260 nm. Se emplearon 3 unidades de absorbancia para cada inmunoprecipitación (30 µg totales). Una alícuota de este sobrenadante se guardó como extracto de cromatina o fracción *"Input"*. A cada una de las alícuotas restantes se les agregó los distintos anticuerpos y que posteriormente fueron incubados toda la noche a 4°C con agitación. Los anticuerpos utilizados fueron: anti RUNX1 (Active Motif), anti p300 (Abcam 14986), anti pCAF (Abcam12188), anti CBP (Abcam 2832). Los complejos antígeno-anticuerpo fueron recogidos con bolitas de Proteína A-Sepharosa CL-4B (Pharmacia) previamente bloqueadas con 100 µg de ADN de salmón fragmentado y 1 mg/mL de BSA y se incubaron en agitación por 1 hora a 4°C. Las bolitas se recogieron por centrifugación a 10000xg por 20 segundos a 4°C. Los complejos antígeno-anticuerpo (pellet) se lavaron 2 veces con buffer de sonicación, 2 veces con buffer de sonicación más 500 mM de NaCl, 2 veces con tampón LiCl (20 mM Tris pH 8.1; 1 mM de EDTA, 250 mM de LiCl, 0.5% v/v NP-40, 0.5% v/v Na deoxicolato) y finalmente 2 veces con buffer Tris-EDTA (Tris 10 mM, EDTA 1mM). Los complejos fueron eluidos con buffer de elución (50 mM de Tris pH 8.0, 1mM de EDTA y 1% v/v de SDS) por 15 minutos a 65°C. Posteriormente, se centrifugaron a velocidad máxima por 20 segundos y al

sobrenadante se agregó NaCl a una concentración final de 200 mM. Finalmente se revertió el cross-linking a 65°C toda la noche en presencia de RNasa A (concentración final 0.1 µg/µL). Después se trataron las muestras con Proteinasa K (concentración final 0.25 µg/µL) por 2 horas a 50°C, se purificó el ADN por extracción fenólica. Los resultados se analizaron por PCR en tiempo real. Para los ensayos de ReChIP, los complejos antígeno-anticuerpo fueron eluidos en 25 µl de DTT 10mM e incubados 30 minutos a 37°C, posteriormente se diluyó 20 veces con buffer de sonicación, se inmunoprecipitó con el nuevo anticuerpo, procediendo a realizar el protocolo detallado anteriormente.

Purificación de ADN desde ChIP

Para extraer y purificar ADN sonificado proveniente de la Inmunoprecipitación de Cromatina se realizó una extracción fenólica, con igual volumen de mezcla fenol-cloroformo (1:1) y posteriormente con cloroformo alcohol-isoamílico (24:1). Entre cada extracción las muestras se centrifugaron a 16000xg por 5 minutos a temperatura ambiente. Luego el ADN purificado se precipitó con 2.5 volúmenes de etanol 100% en presencia de 0.1 volumen de Acetato de Sodio 3 M pH 5.2 toda la noche a -20°C. Posteriormente, las muestras se centrifugaron por 30 minutos a 16000 x g a 4°C, se realizó dos lavados con etanol 70% y finalmente el precipitado se resuspendió en Tris 10 mM pH 8.0.

PCR en tiempo real.

Se utilizó el kit SureStart PCR Master Mix de Stratagene, para ello se utilizaron distintos set de oligonucleótidos, localizados a lo largo del intrón 5 y detallados en la tabla 10. El equipo utilizado para realizar todas las determinaciones corresponde al termociclador en tiempo real MX3000p (Stratagene).



RESULTADOS

Objetivo Específico N1: “Determinar el enriquecimiento de H3K27ac, H3K4me1, H3K4me3, H2A.Z a lo largo del locus del gen *RUNX1*”

A nivel genómico, la mayor proporción de las regiones promotoras se encuentran enriquecidas en H3K27ac.

La acetilación de la lisina 27 de la histona H3 (H3K27ac) ha sido frecuentemente identificada en regiones altamente accesibles de la cromatina. Concordantemente, estudios de secuenciación masiva han evidenciado su enriquecimiento en regiones *cis* regulatorias transcripcionalmente activas (Heintzman, Stuart et al. 2007, Heintzman, Hon et al. 2009).

Tomando en cuenta estos antecedentes, se determinó el enriquecimiento genómico de H3K27ac en células de origen hematopoyético. Para ello, se evaluó data obtenida mediante análisis de ChIPseq en la línea celular GM12878. La data fue descargada desde el servidor ENCODE, desde la biblioteca de datos “Regulation” del Histone Broad Institute. El análisis total de la librería (que se denominó librería de H3K27ac) permitió identificar un total de 56.069 peaks, que corresponden al total de regiones enriquecidas con esta modificación. Con el objetivo de caracterizar la distribución genómica de estas regiones se desarrollaron dos librerías mediante el lenguaje de programación Shell, una de ellas contiene las coordenadas genómicas de inicio y término de todos los genes conocidos a la fecha (librería de regiones intragénicas), mientras que la segunda

librería incluye las coordenadas genómicas de regiones 1.000 pb corriente arriba al primer exón descrito para un gen (librería de regiones promotoras). Ambas librerías fueron integradas como templado al servidor de herramientas GALAXY para su posterior intersección con la librería de H3K27ac.

Del análisis de los datos, determinamos que del total de las regiones mapeadas (56.069), 20.180 (36%) de los peaks alinearon con la librería de regiones promotoras, por ende corresponden a regiones promotoras enriquecidas en H3K27ac. Además, 25.698 (45.8% del total de los peaks) se asocian a regiones intragénicas y 10.191 regiones (18.2%) pertenecientes a la librería de H3K27ac, no intersectan con ninguna de las librerías utilizadas, por ende estos peaks se clasificaron como localizados en regiones intergénicas (**Figura 2A – Tabla 2**). Con el objetivo de caracterizar la localización específica de las regiones intragénicas se utilizó una estrategia similar a la descrita anteriormente donde se desarrolló una librería que contiene las coordenadas genómicas de inicio y término de todos los exones del genoma humano. Posteriormente, esta librería fue intersectada con la librería de localización de peaks presentes en regiones intragénicas. Aquellos peaks que no intersectaron con la librería de exones, fueron clasificados como unidos a regiones intrónicas. Nuestros resultados muestran que del total de los peaks evaluados (25.698), solo el 19.1% de ellos intersectó con la librería de exones (4.900 peaks). Interesantemente, el 80.9% (20.798) de los peaks se localizan en regiones intrónicas (**Figura 2B - Tabla 2**). Considerando que el tamaño del genoma humano (Hg19) esta formado por aproximadamente 3.200 Mb y debido a que en

Tabla 2: Caracterización genómica de las regiones enriquecidas en H3K27ac

Enriquecimiento H3K27ac	Regiones	%
Total de peaks	56.069	100,0
Regiones Promotoras	20.180	36,0
Regiones Intragénicas	25.698	45,8
Regiones Intergénicas	10.191	18,2
Regiones Intrónicas	20.798	80,9
Regiones Exónicas	4.900	19,1

Fuente: Elaboración propia

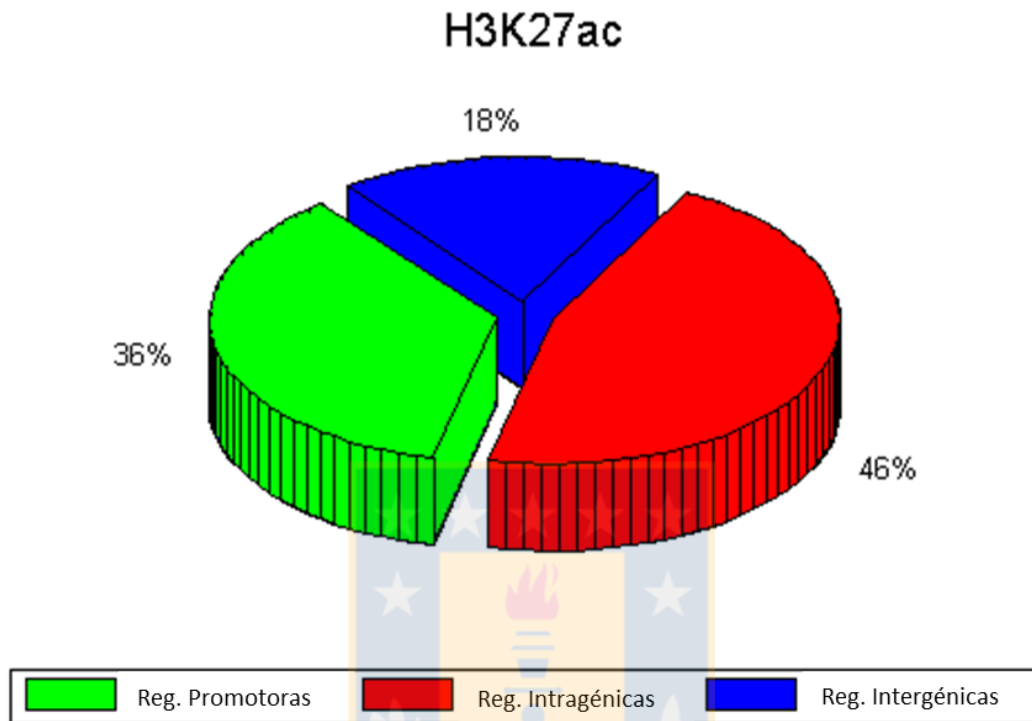
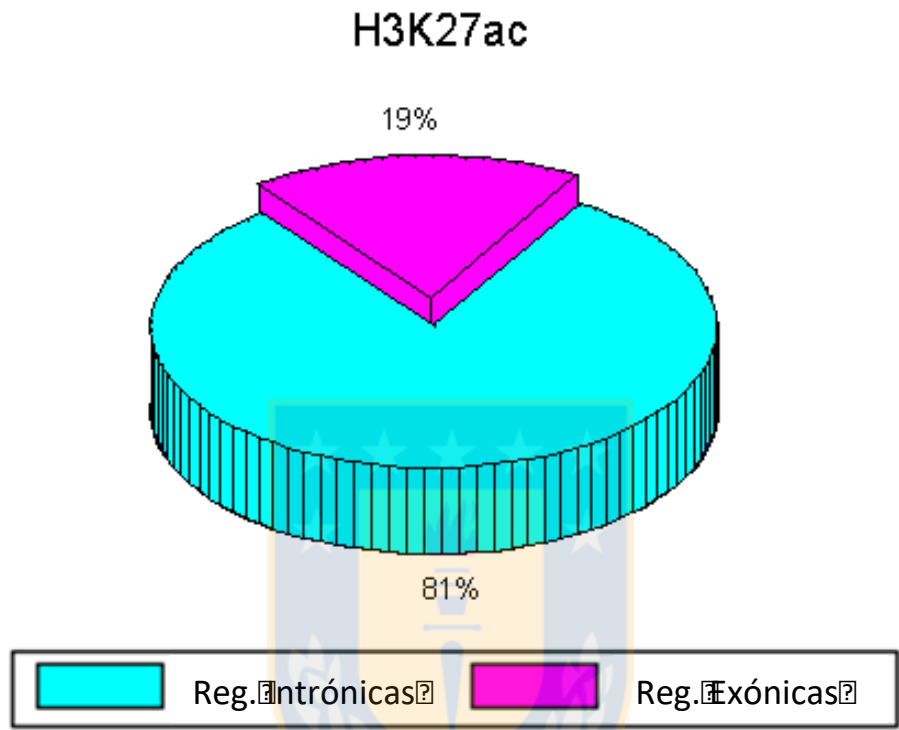


Figura 2A: Regiones intragénicas presentan enriquecimiento en H3K27ac.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K27ac obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell, 2 librerías de datos: 1) La librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha. 2) Librería de Promotores, que contiene las coordenadas genómicas comprendidas desde 1.000 pb corriente arriba del primer exón hasta el nucleótido +1; las librerías fueron intersectadas con los datos de ChIPseq en el servidor GALAXY utilizando el programa “Intersect” del sistema de herramientas BedTools. Las regiones que no intersectaron con ninguna de las librerías utilizadas se denominaron Regiones Intergénicas.



Fuente: Elaboración propia

Figura 2B: Regiones intrónicas presentan un mayor enriquecimiento en H3K27ac.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K27ac obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell la librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha; esta librería se intersectó con los datos de ChIPseq obtenidos en el servidor GALAXY utilizando el programa "Intersect" del sistema de herramientas BedTools. Posteriormente, las regiones identificadas, fueron intersectadas con una nueva librería de coordenadas denominada librería de regiones exónicas, que contiene las coordenadas de inicio y término del exoma. Las regiones identificadas en esta intersección fueron catalogadas como Regiones Exónicas enriquecidas en H3K27ac. Aquellas regiones que no fueron intersectadas se catalogaron como Regiones Intrónicas enriquecidas en H3K27ac.

este estudio el número de genes utilizados en fue de 28.705, para nuestros análisis los promotores representan a nivel genómico son 28.705.000 pares de bases, es decir, menos del 1% del genoma ya que para las regiones promotoras se definió un tamaño estándar de 1.000 pb. Por consiguiente, este escaso número de pares de bases que representan a los promotores, resulta en que se encuentren, como número absoluto, un bajo número de peaks asociados a estas regiones. Esto también es válido para el resto de las regiones, donde si analizamos solamente la distribución de los peaks, estos pueden estar sobrerrepresentados o bien subestimados a causa del valor en pares de bases asociadas a cada una de las regiones definidas (promotoras, intergénicas, exónicas, intrónicas). Por ejemplo, a lo largo del genoma las regiones intergénicas e intrónicas representan una mayor proporción del genoma y por ende, es más probable encontrar en estas un mayor número de peaks solamente por el número de pares de bases presentes. Por este motivo, y con el objetivo de analizar la abundancia real de los peaks en cada una de las regiones estudiadas independientemente de su abundancia en pares de bases a nivel genómico, decidimos normalizar el contenido de peaks en función al porcentaje de representación genómica de regiones promotoras, intrónicas, exónicas e intergénicas. Para ello, obtuvimos el cociente entre el número de peaks y el porcentaje de representación genómica de cada una de estas regiones (**Tabla 3**). Al evaluar esta normalización podemos apreciar que la mayor parte de los peaks se asocia preferentemente a regiones promotoras y exónicas, y solo un bajo número del total de regiones intrónicas e intergénicas se encuentra enriquecido en H3K27ac (**Figura 3**).

Tabla 3: Normalización genómica del enriquecimiento en H3K27ac

	Tamaño genómico (pb)	Representación Genómica :RG (%)	Total peaks mapeados	Normalización Enriquecimiento Genómico (peaks/% RG)
Regiones Promotoras	28.705.000	0,9	20.180	22.422
Regiones Intergénicas	1.962.603.705	60,8	10.191	168
Regiones Intrónicas	1.196.281.121	37,0	20.798	562
Regiones Exónicas	42.364.894	1,3	4.900	3.769
Totales				
Genoma Hg19	3.229.954.720	100	56.069	

Fuente: Elaboración propia

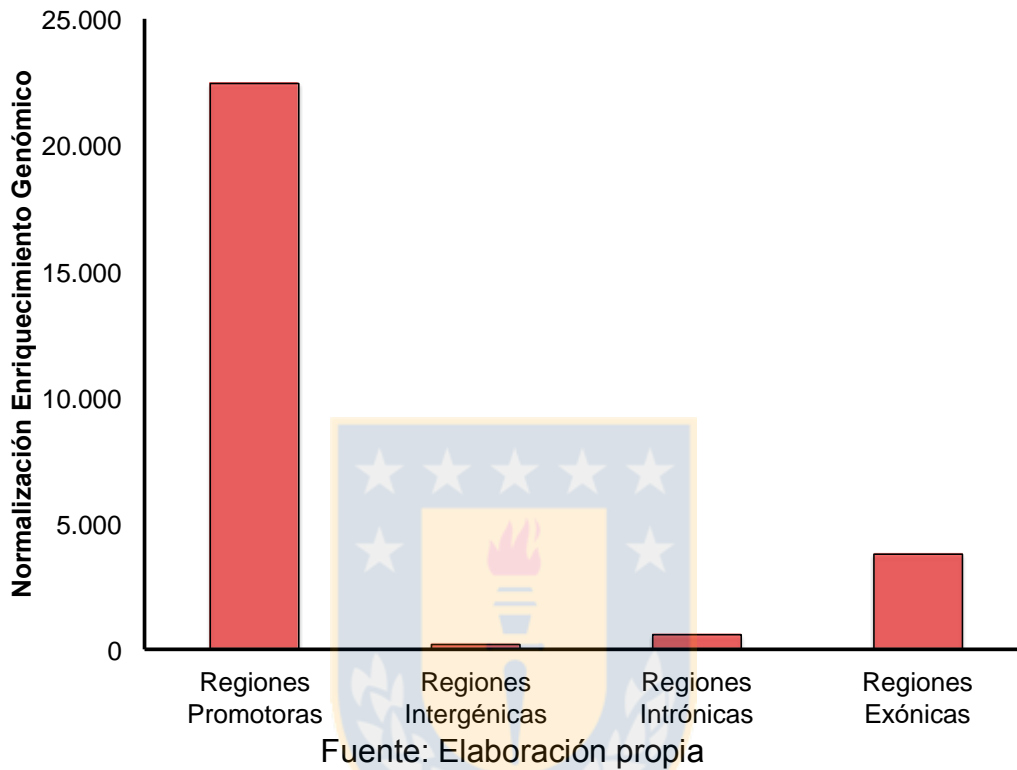


Figura 3: La mayor proporción de las regiones promotoras, se encuentran efectivamente enriquecidas en H3K27ac.

Utilizando la caracterización de cobertura genómica resuelta por Sakharkar y cols (Sakharkar, Perumal et al. 2005), se obtuvo con respecto al genoma completo (Hg19) la totalidad de pares de bases que representan a las regiones intrónicas, exónicas e intergénicas. La cobertura de las regiones promotoras fue calculada multiplicando el número total de genes por 1.000 pb (longitud utilizada operacionalmente para definir regiones promotoras). Con estos datos se procedió a realizar la normalización de la distribución de peaks mapeados para H3K27ac, dividiendo el número total de peaks identificados para cada librería por el porcentaje de representación genómica de las regiones promotoras, intrónicas, exónicas e intergénicas respectivamente. Los resultados son representados de forma relativa, a modo de “Normalización de Enriquecimiento Genómico”.

A nivel genómico, la mayor proporción de las regiones promotoras y exónicas se encuentran enriquecidas en H3K4me3 y H3K4me1

La metilación de la lisina 4 de la histona H3 (H3K4me) ha sido ligada a la activación transcripcional en una amplia variedad de células eucarióticas (Barski, Cuddapah et al. 2007, Heintzman, Stuart et al. 2007, Eissenberg and Shilatifard 2010). Este residuo de lisina puede ser mono, di o trimetilado, agregando un componente de complejidad adicional a la regulación transcripcional mediada por esta metilación. Análisis recientes han determinado que existe una correlación entre el estado de metilación de H3K4 y el nivel de expresión génica de un gen. Actualmente es prácticamente un consenso que la trimetilación de la lisina 4 de la histona H3 (H3K4me3) se asocia a las regiones promotoras de los genes transcripcionalmente activos y de hecho existe una fuerte correlación entre esta modificación y la tasa de expresión génica (Heintzman, Stuart et al. 2007).

Tomando en consideración estos antecedentes, se evaluó data obtenida mediante análisis de ChIPseq utilizando un anticuerpo anti H3K4me3 en la línea celular linfoblástica GM12878. La data fue descargada desde el servidor ENCODE y analizada en el servidor de herramientas GALAXY. Nuestro análisis identificó un total de 73.952 regiones enriquecidas en esta marca epigenética. Interesantemente, nuestros resultados muestran que solo 12.417 (16.8%) peaks fueron mapeados en regiones promotoras, lo cual contrasta con las 37298 (50,4) y 24.237 (32.8) regiones identificadas en regiones intragénicas e intergénicas respectivamente (**Figura 4A – Tabla 4**). Al realizar el análisis detallado de las regiones intragénicas (37.298), el 42,5% de ellas (15.971 peaks), se mapearon en

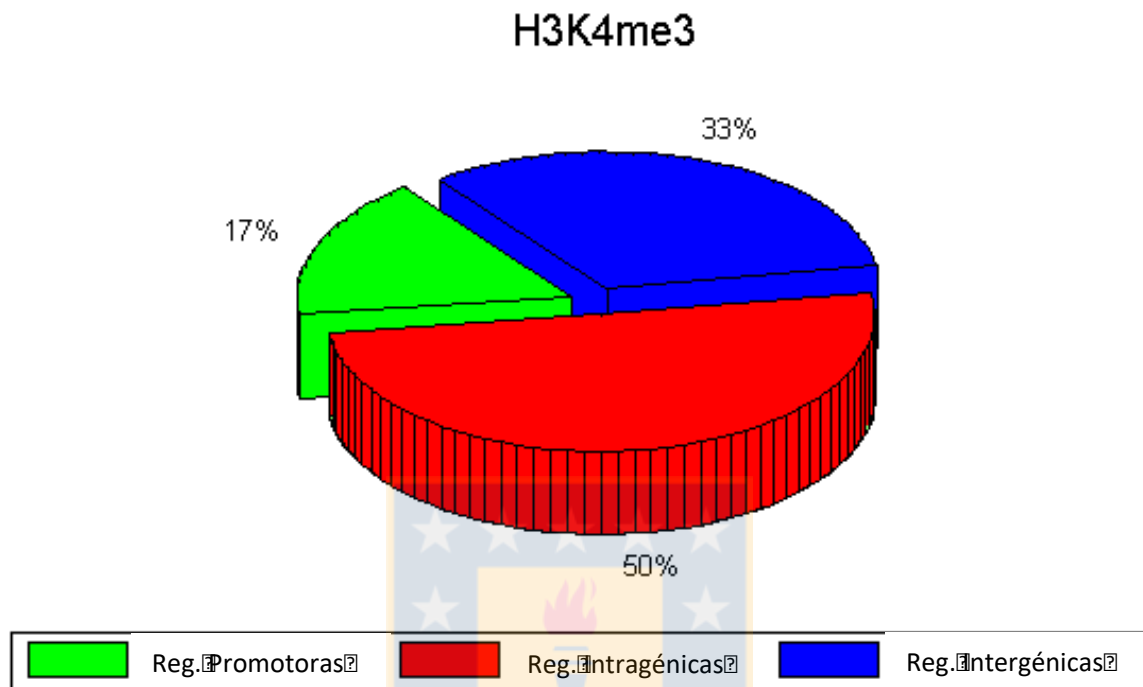
regiones intrónicas y el 57.2% (21.327) a regiones exónicas (**Figura 4B – Tabla 4**). Al determinar el enriquecimiento efectivo de dichas regiones, esto es, tomando en consideración la proporción real de cada una de las regiones analizadas con respecto al genoma (**Tabla 5**), nuestros resultados demuestran que efectivamente, la mayor proporción de las regiones promotoras y exónicas del genoma se encuentran enriquecidas en H3K4me3 y solo un bajo número del total de los intrones y regiones intergénicas, a nivel genómico, presenta dicha modificación (**Figura 5**).

Al evaluar la distribución de la monometilación de H3K4 (H3K4me1), se encontró un mayor número de regiones enriquecidas en esta marca (130.000). De las regiones identificadas, el 56,4% (73.340 peaks) fue mapeada en regiones intergénicas y solo el 11.3% (14.753 peaks) se localizan en regiones promotoras y el 32.2% (41.907 peaks) en regiones intragénicas (**Figura 6A – Tabla 6**), siendo mayoritaria su asociación a regiones intrónicas con un 76,4% de los peaks (32.018 regiones) (**Figura 6B – Tabla 6**). Al igual que en el caso anterior, al estudiar la proporción de regiones promotoras, intrónicas, exónicas e intergénicas enriquecidas en H3K4me1 con respecto al total de ellas a nivel genómico (**Tabla 7**), se encontró que del total de promotores, una alta proporción de ellos, al igual que en las regiones exónicas, se encuentran enriquecidos en dichas marcas. No obstante, solo una reducida proporción de intrones y regiones intergénicas presenta dicha modificación (**Figura 7**).

Tabla 4: Caracterización genómica de las regiones enriquecidas en H3K4me3

Enriquecimiento H3K4me3	Regiones Mapeadas	%
Total de Lecturas	73.952	100,0
Regiones Promotoras	12.417	16,8
Regiones Intragénicas	37.298	50,4
Regiones Intergénicas	24.237	32,8
Regiones Intrónicas	15.971	42,8
Regiones Exónicas	21.327	57,2

Fuente: Elaboración propia

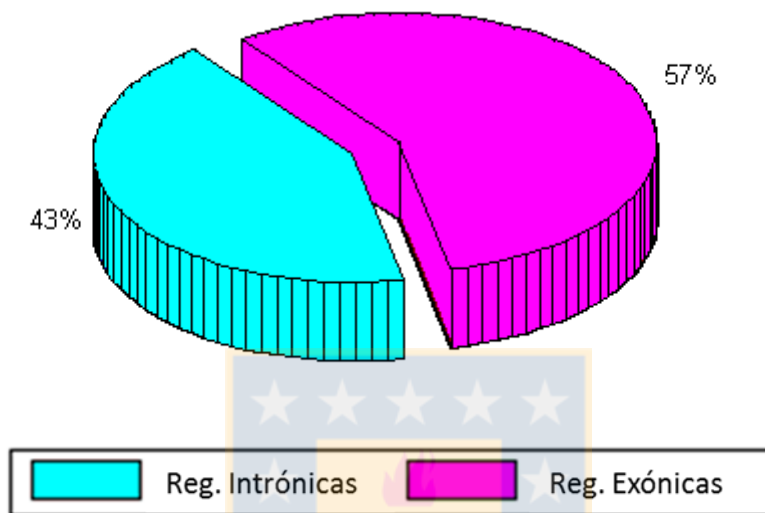


Fuente: Elaboración propia

Figura 4A: A nivel genómico, la mayor parte de las regiones enriquecidas en H3K4me3 se localiza en regiones intragénicas.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K4me3 obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell, 2 librerías de datos: La librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha y la Librería de Promotores, que contiene las coordenadas genómicas comprendidas desde 1.000 pb corriente arriba del primer exón hasta el nucleótido +1. Ambas librerías fueron intersectadas con los datos de ChIPseq en el servidor GALAXY utilizando el programa “Intersect” del sistema de herramientas BedTools. Las regiones que no intersectaron con ninguna de las librerías utilizadas se denominaron Regiones Intergénicas.

H3K4me3



Fuente: Elaboración propia

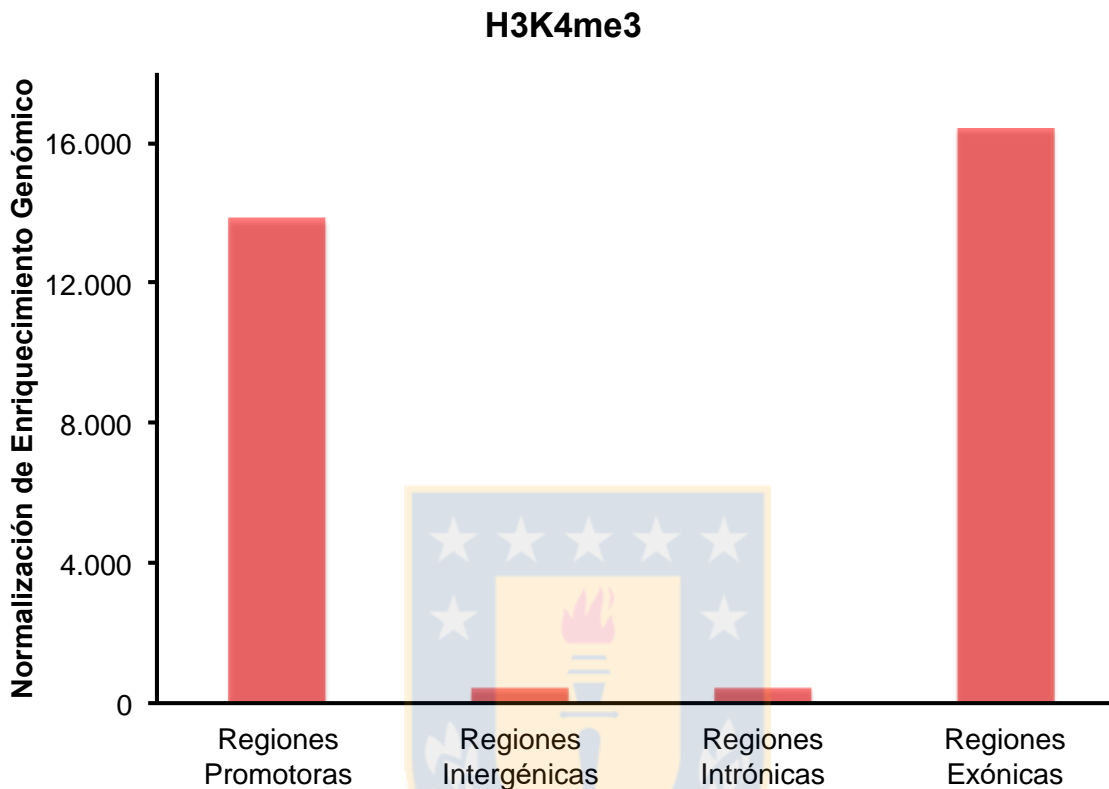
Figura 4B: De las regiones intragénicas enriquecidas en H3K4me3, la mayor parte corresponde a regiones exónicas.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K4me3 obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell La librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha, esta librería se intersectó con los datos de ChIPseq obtenidos en el servidor GALAXY utilizando el programa "Intersect" del sistema de herramientas BedTools. Posteriormente, las regiones identificadas fueron intersectadas con una nueva librería de coordenadas denominada librería de regiones exónicas, que contiene las coordenadas de inicio y término del exoma. Las regiones identificadas en esta intersección fueron catalogadas como Regiones Exónicas enriquecidas en H3K4me3. Aquellas regiones que no fueron intersectadas se catalogaron como Regiones Intrónicas enriquecidas en H3K4me3.

Tabla 5 : Normalización genómica del enriquecimiento en H3K4me3

	Tamaño genómico (pb)	Representación Genómica :RG (%)	Total peaks mapeados	Normalización Enriquecimiento Genómico (peaks/% RG)
Regiones Promotoras	28.705.000	0,9	12.417	13.797
Regiones Intergénicas	1.962.603.705	60,8	24.237	399
Regiones Intrónicas	1.196.281.121	37,0	15.971	432
Regiones Exónicas	42.364.894	1,3	21.327	16.405
Totales				
Genoma Hg19	3.229.954.720	100	73.952	

Fuente: Elaboración propia



Fuente: Elaboración propia

Figura 5: La mayor proporción de las regiones promotoras y exónicas, se encuentran efectivamente enriquecidas en H3K4me3.

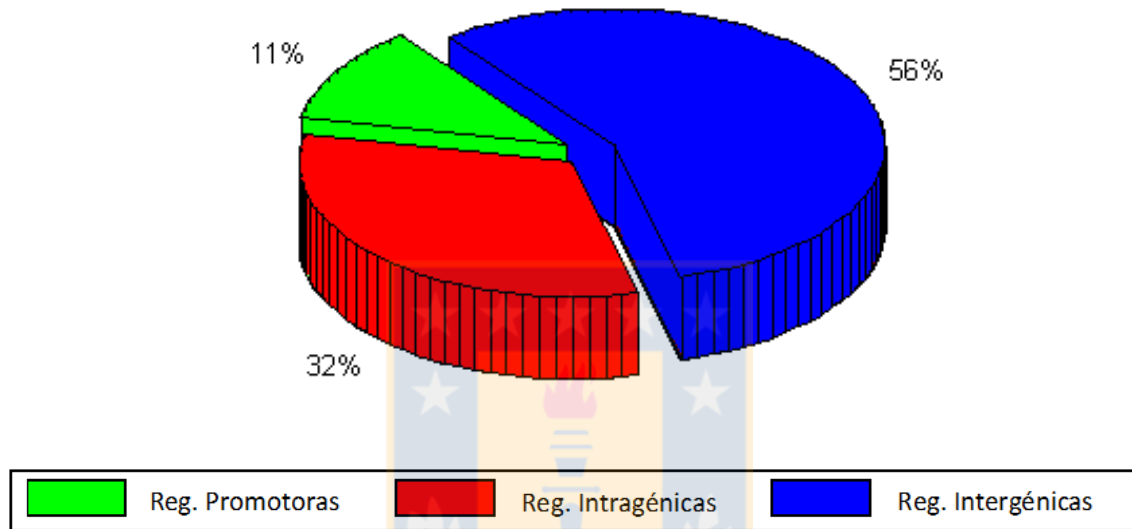
Utilizando la caracterización de cobertura genómica resuelta por Sakharkar y cols (Sakharkar, Perumal et al. 2005), se obtuvo con respecto al genoma completo (Hg19) la totalidad de pares de bases que representan a las regiones intrónicas, exónicas e intergénicas. La cobertura de las regiones promotoras fue calculada multiplicando el número total de genes por 1.000 pb (longitud utilizada operacionalmente para definir regiones promotoras). Con estos datos se procedió a realizar la normalización de la distribución de peaks mapeados para H3K4me3, dividiendo el número total de peaks identificados para cada librería por el porcentaje de representación genómica de las regiones promotoras, intrónicas, exónicas e intergénicas respectivamente. Los resultados son representados de forma relativa, a modo de “Normalización de Enriquecimiento Genómico”.

Tabla 6: Caracterización genómica de las regiones enriquecidas en H3K4me1

Enriquecimiento en H3K4me1	Regiones mapeadas	%
Total de Lecturas	130.000	100,0
Regiones Promotoras	14.753	11,3
Regiones Intragénicas	41.907	32,2
Regiones Intergénicas	73.340	56,4
Regiones Intrónicas	32.018	76,4
Regiones Exónicas	9.889	23,6

Fuente: Elaboración propia

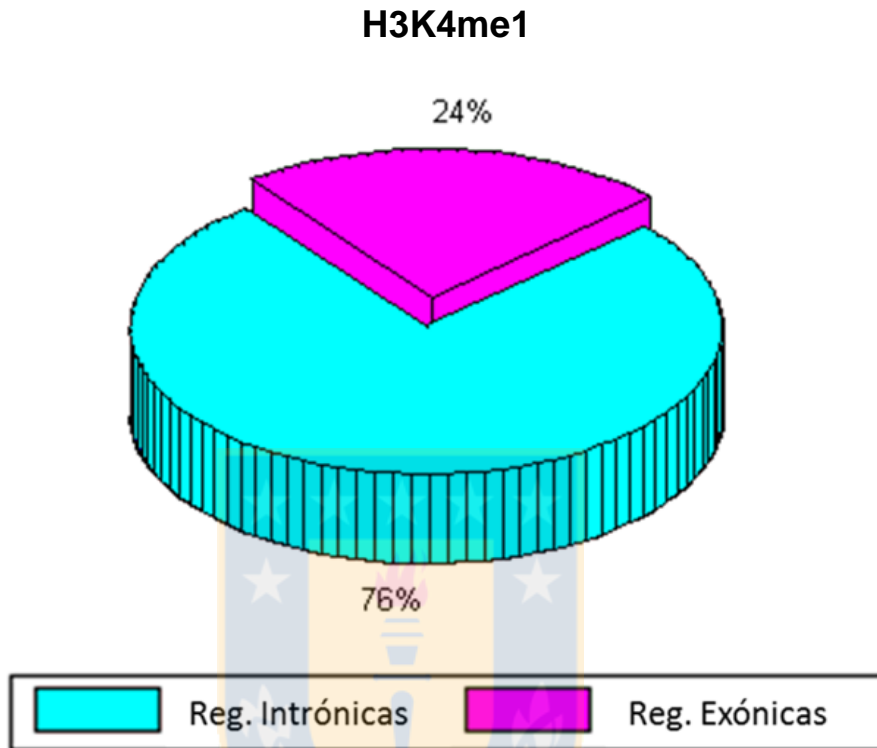
H3K4me1



Fuente: Elaboración propia

Figura 6A: A nivel genómico, la mayor parte de las regiones enriquecidas en H3K4me1 se localiza en regiones intergénicas.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K4me1 obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell, 2 librerías de datos: la Librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha, y la Librería de Promotores, que contiene las coordenadas genómicas comprendidas desde 1.000 pb corriente arriba del primer exón hasta el nucleótido +1. Ambas librerías fueron intersectadas con los datos de ChIPseq en el servidor GALAXY utilizando el programa "Intersect" del sistema de herramientas BedTools. Las regiones que no intersectaron con ninguna de las librerías utilizadas se denominaron Regiones Intergénicas.



Fuente: Elaboración propia

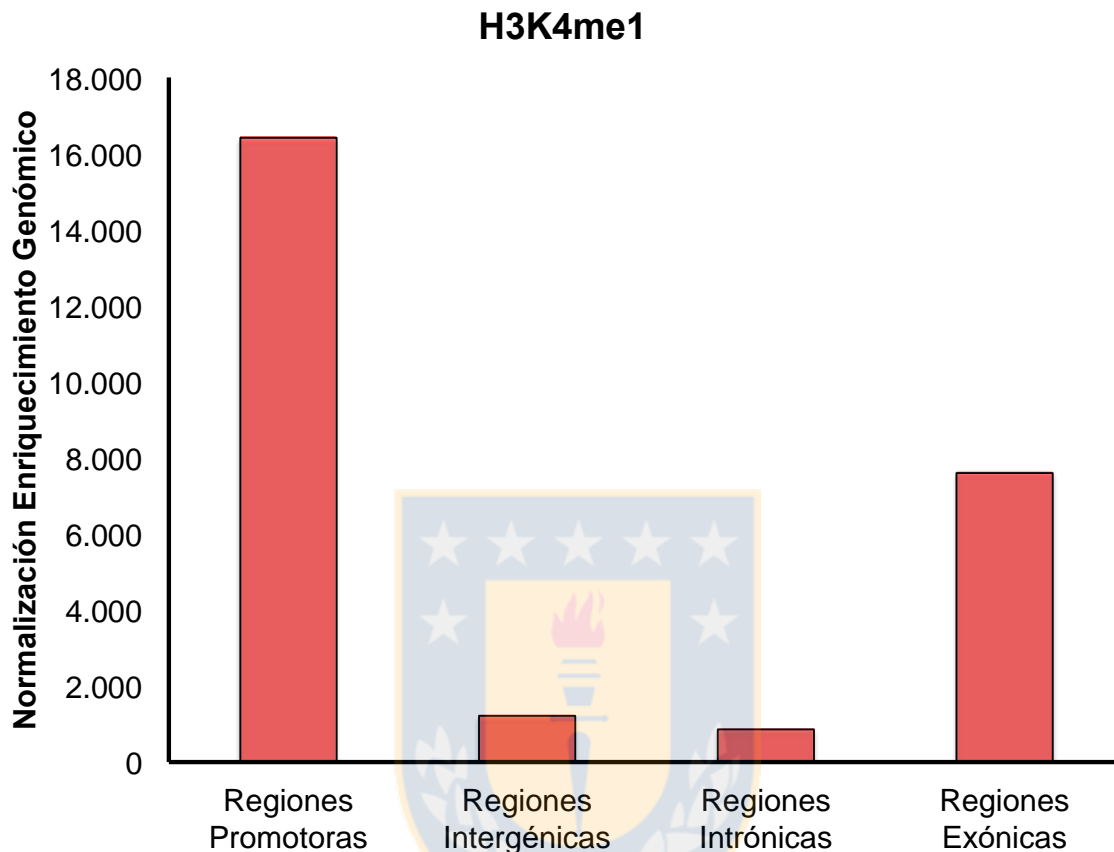
Figura 6B: De las regiones intragénicas enriquecidas en H3K4me1, la mayor parte corresponde a regiones exónicas.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H3K4me1 obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell la Librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha; esta librería se intersectó con los datos de ChIPseq obtenidos en el servidor GALAXY utilizando el programa “*Intersect*” del sistema de herramientas BedTools. Posteriormente las regiones identificadas fueron intersectadas con una nueva librería de coordenadas denominada Librería de Regiones Exónicas, que contiene las coordenadas de inicio y término del exoma. Las regiones identificadas en esta intersección fueron catalogadas como Regiones Exónicas enriquecidas en H3K4me1. Aquellas regiones que no fueron intersectadas se catalogaron como Regiones Intrónicas enriquecidas en H3K4me1.

Tabla 7: Normalización genómica del enriquecimiento en H3K4me1

	Tamaño genómico (pb)	Representación Genómica :RG (%)	Total peaks mapeados	Normalización Enriquecimiento Genómico (peaks/% RG)
Regiones Promotoras	28.705.000	0,9	14.753	16.392
Regiones Intergénicas	1.962.603.705	60,8	73.340	1.206
Regiones Intrónicas	1.196.281.121	37,0	32.018	865
Regiones Exónicas	42.364.894	1,3	9.889	7.607
Totales				
Genoma Hg19	3.229.954.720	100	130.000	

Fuente: Elaboración propia



Fuente: Elaboración propia

Figura 7: La mayor proporción de las regiones promotoras y exónicas, se encuentran efectivamente enriquecidas en H3K4me1.

Utilizando la caracterización de cobertura genómica resuelta por Sakharkar y cols (Sakharkar, Perumal et al. 2005), se obtuvo con respecto al genoma completo (Hg19) la totalidad de pares de bases que representan a las regiones intrónicas, exónicas e intergénicas. La cobertura de las regiones promotoras fue calculada multiplicando el número total de genes por 1.000 pb (longitud utilizada peracionalmente para definir regiones promotoras). Con estos datos se procedió a realizar la normalización de la distribución de peaks mapeados para H3K4me1 dividiendo el número total de peaks identificados para cada librería por el porcentaje de representación genómica de las regiones promotoras, intrónicas, exónicas e intergénicas, respectivamente. Los resultados son representados de forma relativa, a modo de “Normalización de Enriquecimiento Genómico”.

A nivel genómico, la mayor parte de las regiones enriquecidas en H2A.Z se localiza en regiones promotoras

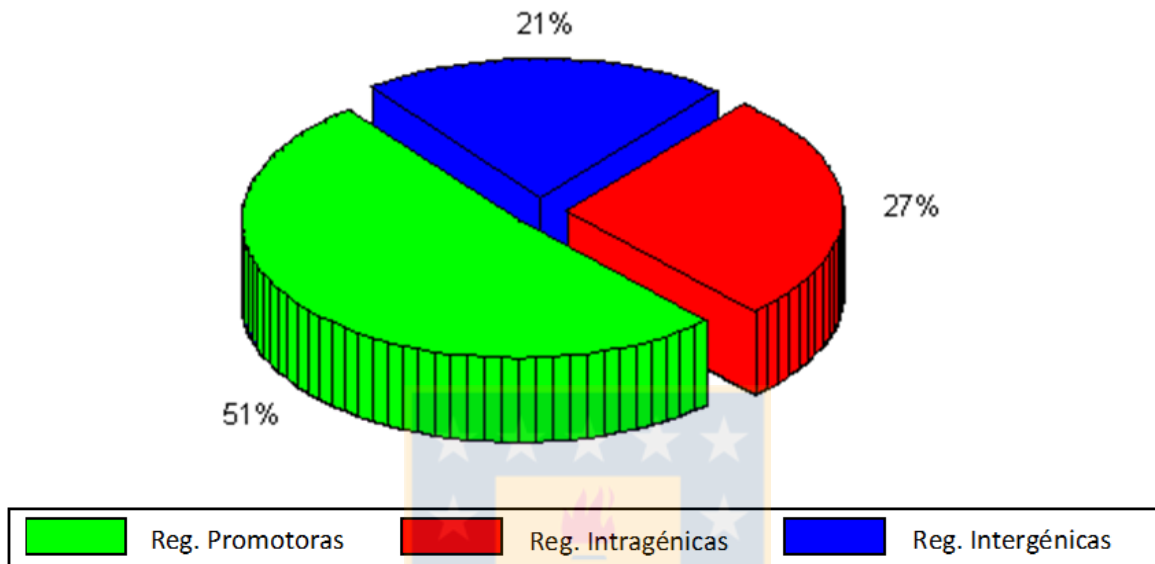
La presencia de la variante de histonas H2A.Z ha sido ampliamente reportada asociada a regiones promotoras (Li, Pattenden et al. 2005, Soboleva, Nekrasov et al. 2011, Nekrasov, Amrichova et al. 2012, Nekrasov, Soboleva et al. 2013, Rosenbloom, Armstrong et al. 2015), por lo que se decidió evaluar su distribución genómica en células GM12878. Al igual que lo descrito anteriormente, se evaluó data obtenida mediante análisis de ChIPseq utilizando un anticuerpo anti H2A.Z en la línea celular GM12878. La data fue descargada desde el servidor ENCODE (Rosenbloom, Armstrong et al. 2015), la posterior caracterización de las regiones fue realizada en el servidor GALAXY (Blankenberg and Hillman-Jackson 2014). Del análisis de los resultados se determinó que H2A.Z se asocia genómicamente a un total de 95.358 regiones. De estas, y de acuerdo a lo esperado, la gran mayoría corresponde a regiones promotoras con un 51.5% del total de los peaks (49.086). También encontramos una asociación importante a regiones intragénicas e intergénicas, con un 27% (25.774) y 21.5% (20.498) de asociaciones respectivamente (**Figura 8A – Tabla 8**). Sorprendentemente, del total de regiones intragénicas encontradas, el 76% de los peaks (19.653 regiones) se localizan en intrones y solo un 23.7% (6.121 regiones) en exones (**Figura 8B – Tabla 8**). No obstante, al evaluar la proporción total de regiones promotoras, intrónicas, exónicas e intergénicas, con respecto al total de ellas en el genoma, enriquecidas en H2A.Z (**Tabla 9**) se evidenció que, de todas las regiones

Tabla 8: Caracterización genómica de las regiones enriquecidas en H2A.Z

Enriquecimiento H2A.Z	Regiones mapeadas	%
Total de Lecturas	95.358	100
Regiones Promotoras	49.086	51,5
Regiones Intragénicas	25.774	27,0
Regiones Intergénicas	20.498	21,5
Regiones Intrónicas	19.653	76,3
Regiones Exónicas	6.121	23,7

Fuente: Elaboración propia

H2A.Z



Fuente: Elaboración propia

Figura 8A: A nivel genómico, la mayor parte de las regiones enriquecidas en H2A.Z se localiza en regiones promotoras.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H2A.Z obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell, 2 librerías de datos: La Librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha y la Librería de Promotores, que contiene las coordenadas genómicas comprendidas desde 1.000 pb corriente arriba del primer exón hasta el nucleótido +1. Ambas librerías fueron intersectadas con los datos de ChIPseq en el servidor GALAXY utilizando el programa "Intersect" del sistema de herramientas BedTools. Las regiones que no intersectaron con ninguna de las librerías utilizadas se denominaron Regiones Intergénicas.

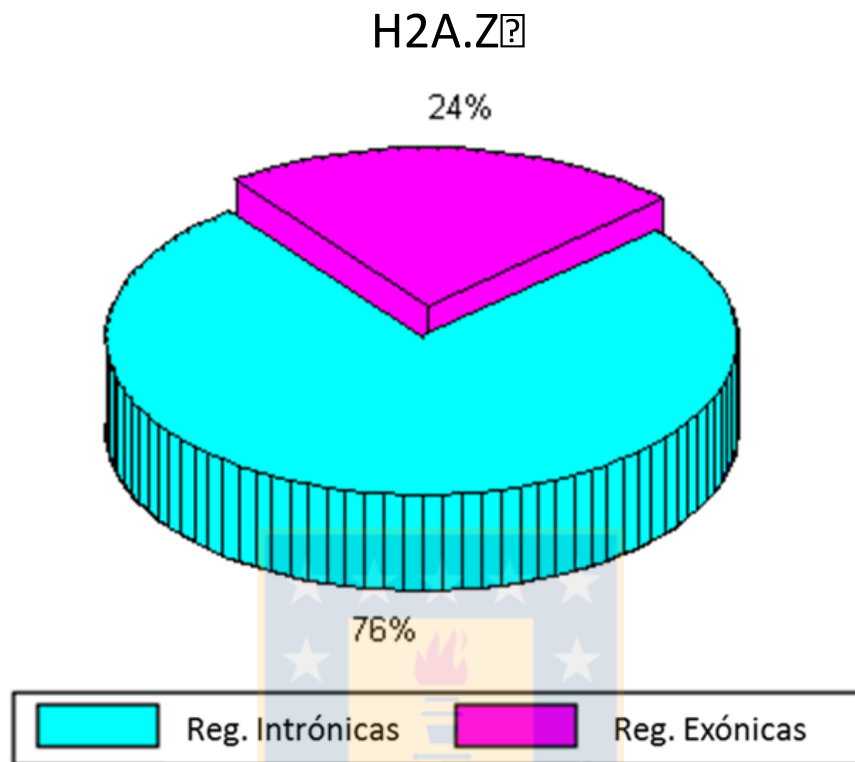


Figura 8B: De las regiones intragénicas enriquecidas en H2A.Z, la mayor parte corresponde a intrones.

Desde el servidor ENCODE se descargó y analizó data de ChIPseq para la modificación de histona H2A.Z obtenida desde células GM12878. Simultáneamente se desarrolló con lenguaje de programación Shell la Librería Intragénica, con las coordenadas de inicio y término de todos los genes conocidos a la fecha, esta librería se intersectó con los datos de ChIPseq obtenidos, en el servidor GALAXY utilizando el programa “*Intersect*” del sistema de herramientas BedTools. Posteriormente, las regiones identificadas fueron intersectadas con una nueva librería de coordenadas denominada Librería de Regiones Exónicas, que contiene las coordenadas de inicio y término del exoma. Las regiones identificadas en esta intersección fueron catalogadas como Regiones Exónicas Enriquecidas en H2A.Z. Aquellas regiones que no fueron intersectadas se catalogaron como Regiones Intrónicas Enriquecidas en H2A.Z.

Tabla 9: Normalización genómica del enriquecimiento en H2A.Z

	Tamaño genómico (pb)	Representación Genómica :RG (%)	Total peaks mapeados	Normalización Enriquecimiento Genómico (peaks/% RG)
Regiones Promotoras	28.705.000	0,9	49.086	54.540
Regiones Intergénicas	1.962.603.705	60,8	20.498	337
Regiones Intrónicas	1.196.281.121	37,0	19.653	531
Regiones Exónicas	42.364.894	1,3	6.121	4.708
Totales				
Genoma Hg19	3.229.954.720	100	95.358	

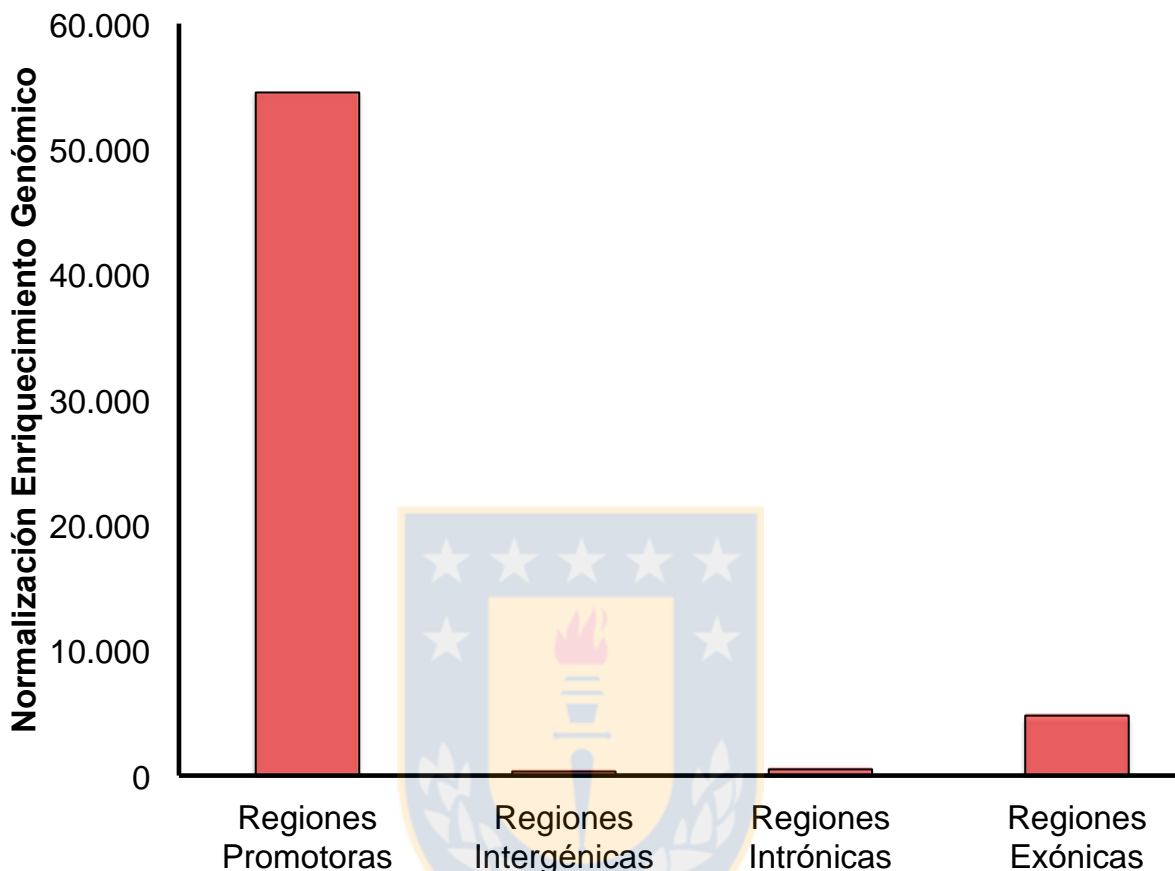
Fuente: Elaboración propia

evaluadas, las regiones promotoras, son las regiones que se encuentran mayormente enriquecidas, con respecto al total de ellas en el genoma. De igual modo, llama la atención la identificación de un reducido número de regiones intrónicas, exónicas e intergénicas enriquecidas en esta variante de histonas **(Figura 9)**.

El locus del gen RUNX1 presenta múltiples regiones enriquecidas en marcas epigenéticas asociadas a módulos cis reguladores

El gen *RUNX1* es blanco frecuente de translocaciones cromosómicas. A la fecha, análisis moleculares han revelado que a pesar del gran tamaño del locus de *RUNX1* (260kb), los sitios de quiebre cromosomal se localizan en intrones específicos para las distintas translocaciones que afectan a este gen. Por ejemplo, se ha establecido que los sitios de quiebre para la t(8;21), una de las translocaciones cromosómicas más recurrentes presentes en pacientes con leucemia mieloide aguda, han sido mapeados en el intrón 5 del gen *RUNX1*, mientras que en la translocación (12; 21) los quiebres se localizan en el intrón 1 y en los intrones 6 y 7 para la t(3;21). Además, al analizar la estructura de cromatina de estas regiones, se ha evidenciado hiperacetilación de histonas H3 y H4 y la presencia de sitios de hipersensibilidad a DNasa I, lo que sugiere que en estas regiones el ADN se encuentra altamente descompactado. A la fecha, la hipersensibilidad a DNasa I es una de las características más utilizada en la identificación de nuevos elementos reguladores. Esto se debe a que la hipersensibilidad a esta enzima se relaciona con una estructura de cromatina

H2A.Z



Fuente: Elaboración propia

Figura 9: La mayor proporción de las regiones promotoras, se encuentran efectivamente enriquecidas en H2A.Z.

Utilizando la caracterización de cobertura genómica resuelta por Sakharkar y cols (Sakharkar, Perumal et al. 2005), se obtuvo con respecto al genoma completo (Hg19) la totalidad de pares de bases que representan a las regiones intrónicas, exónicas e intergénicas. La cobertura de las regiones promotoras fue calculada multiplicando el número total de genes por 1.000 pb (longitud utilizada operacionalmente para definir regiones promotoras). Con estos datos se procedió a realizar la normalización de la distribución de peaks mapeados para H2A.Z, dividiendo el número total de peaks identificados para cada librería por el porcentaje de representación genómica de las regiones promotoras, intrónicas, exónicas e intergénicas, respectivamente. Los resultados son representados de forma relativa, a modo de “Normalización de Enriquecimiento Genómico”.

altamente accesible a la maquinaria transcripcional característica propia de regiones *cis* transcripcionalmente activas. De hecho, ha sido ampliamente reportada la presencia de un peak de hipersensibilidad a DNasa I tanto en regiones promotoras como en enhancers. Lo cual podría sugerir la presencia de elementos *cis* reguladores presentes en las zonas de quiebres descritas para el gen *RUNX1*.

Considerando que nuestros análisis anteriores revelaron que, de la totalidad de intrones del genoma, solo unas cuantas regiones intrónicas presentan enriquecimiento en las marcas epigenéticas asociadas a regiones *cis* reguladoras (H3K27ac, H3K4me3, H3K4me1 y H2A.Z), decidimos analizar en detalle el enriquecimiento de estas marcas epigenéticas en el locus de *RUNX1*, con especial énfasis en los intrones de este gen, para de identificar potenciales módulos *cis* regulatorios en regiones de quiebre cromosomal. Para ello se visualizó la data anteriormente estudiada, en el visor del servidor Genome Browser del proyecto ENCODE y se utilizó una ventana genómica de 260 kb (el tamaño total del locus de *RUNX1*).

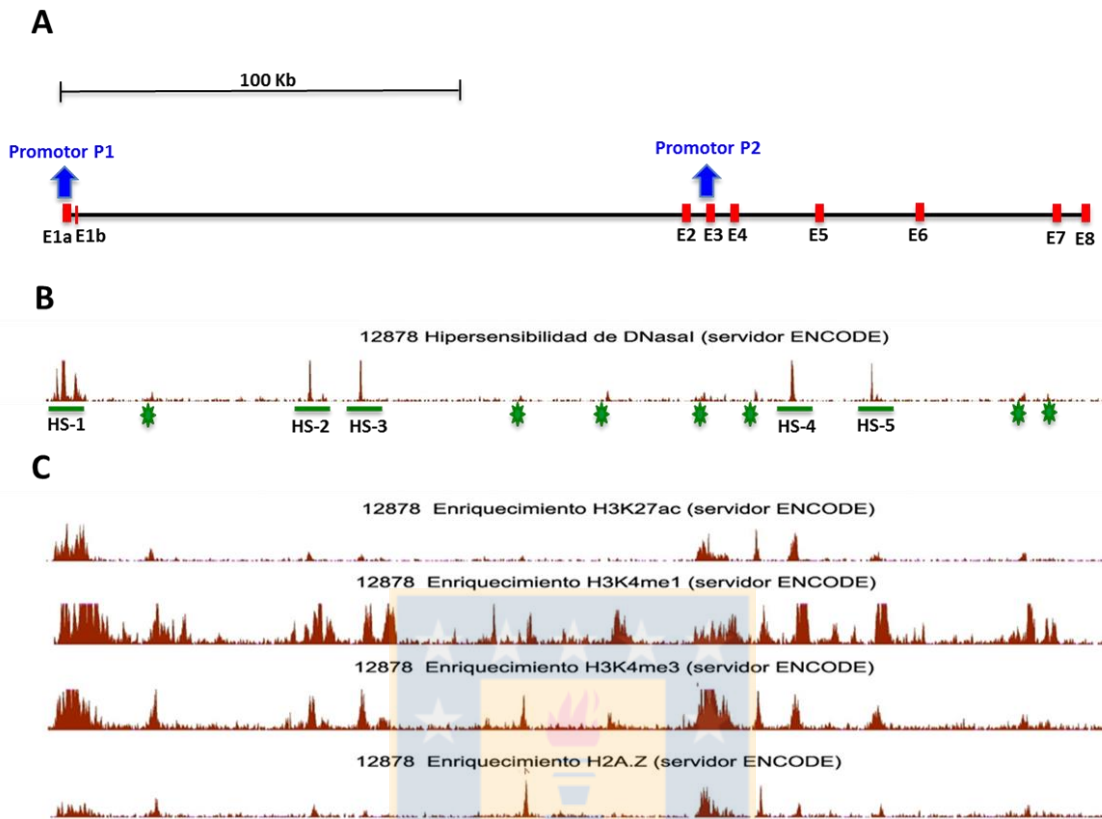
Tomando en cuenta que la hipersensibilidad a DNasa I es una herramienta útil en la identificación de regiones *cis* transcripcionalmente activas, se agregó a nuestro análisis data obtenida desde el servidor ENCODE de ensayos de hipersensibilidad a DNasa I acoplada a secuenciación masiva en la línea celular linfoblástica GM12878. Considerando que la accesibilidad a la cromatina es uno de los principales determinantes en el proceso de transcripción génica, primeramente identificamos regiones del gen *RUNX1* que presentan un alto grado de Hipersensibilidad a DNasa I.

Nuestros análisis nos muestran que a lo largo del locus existen 5 regiones altamente Hipersensibles a DNasa I las que fueron designadas del 1-5. Estas regiones fueron mapeadas co-localizando con el promotor P1 del gen *RUNX1* y con los intrones 1, 4 y 5. Si bien es posible apreciar cierto grado de hipersensibilidad en algunas otras regiones, las que son destacadas en la figura con un asterisco de color rojo (**Figura 10 panel B**), la hipersensibilidad presente no es tan marcada como las regiones previamente descritas.

Al observar la distribución en el locus de la H3K27ac y H2A.Z, podemos determinar que las 5 regiones hipersensibles a DNasa I presentan a su vez distintos grados de enriquecimiento a H3K27ac y H2A.Z. Al evaluar el enriquecimiento en H3K4me1 y H3K4me3 podemos apreciar que existe un gran número de regiones altamente enriquecidas en una o ambas marcas, sin embargo al acotar nuestra búsqueda a las regiones hipersensibles podemos observar que todas estas presentan un alto grado de enriquecimiento en H3K4me3 y H3K4me1 (**Figura 10 panel C**).

Es importante destacar que, si bien al analizar la data en su conjunto, a primera vista destacan principalmente dos regiones: la región 1 que presenta enriquecimiento en todas las marcas analizadas y que se localiza sobrelapando el exón 1- región 5' del intrón 1 y una segunda región, con una menor Hipersensibilidad a DNasa I, localizada entre el exón 2 y el exón 4, destacadas con una flecha azul y denominadas Promotor P1 y P2 (en letras azules), en el esquema del locus (**Figura 10A y Figura 11**), estas regiones se sobreponen con los promotores P1 y P2 descritos para el gen *RUNX1*.

Locus de *RUNX1*



Fuente: Elaboración propia

Figura 10: El locus del gen *RUNX1* presenta marcas epigenéticas asociadas a módulos *cis* reguladores.

Los datos analizados corresponden a datos de ChIPseq obtenidos a partir de la línea GM12878 y utilizando anticuerpos contra las modificaciones de histonas H3K27ac, H3K4me1, H3K4me3, la variante de histonas H2A.Z y ensayos de hipersensibilidad a DNasa I la que esta libremente accesible y descargable desde el servidor ENCODE. Todos los datos fueron visualizados utilizando la herramienta Genome Browser del servidor ENCODE con una ventana genómica de 260 Kb. En el panel A, se muestra una representación esquemática del locus de *RUNX1*, en líneas verticales rojas se resaltan las regiones exónicas (E1- E8), la línea horizontal de color negro entre los exones representa a las regiones intrónicas del gen. Además, señalizado con flechas de color azul, y sobre los exones E1 y E3, se representan los promotores alternativos P1 y P2, respectivamente, descritos para el gen *RUNX1*. En el panel B se representan los datos obtenidos mediante ensayos de Hipersensibilidad a DNasa I (HS1-5: resaltados también con una línea sólida de color verde bajo el peak correspondiente), además se destacan con asterisco verde las regiones con menor grado de hipersensibilidad presentes. En el panel C se presenta los datos obtenidos mediante ChIPseq para las modificaciones de histonas H3K27ac, H3K4me1, H3K4me3 y la variante de histonas H2A.Z.

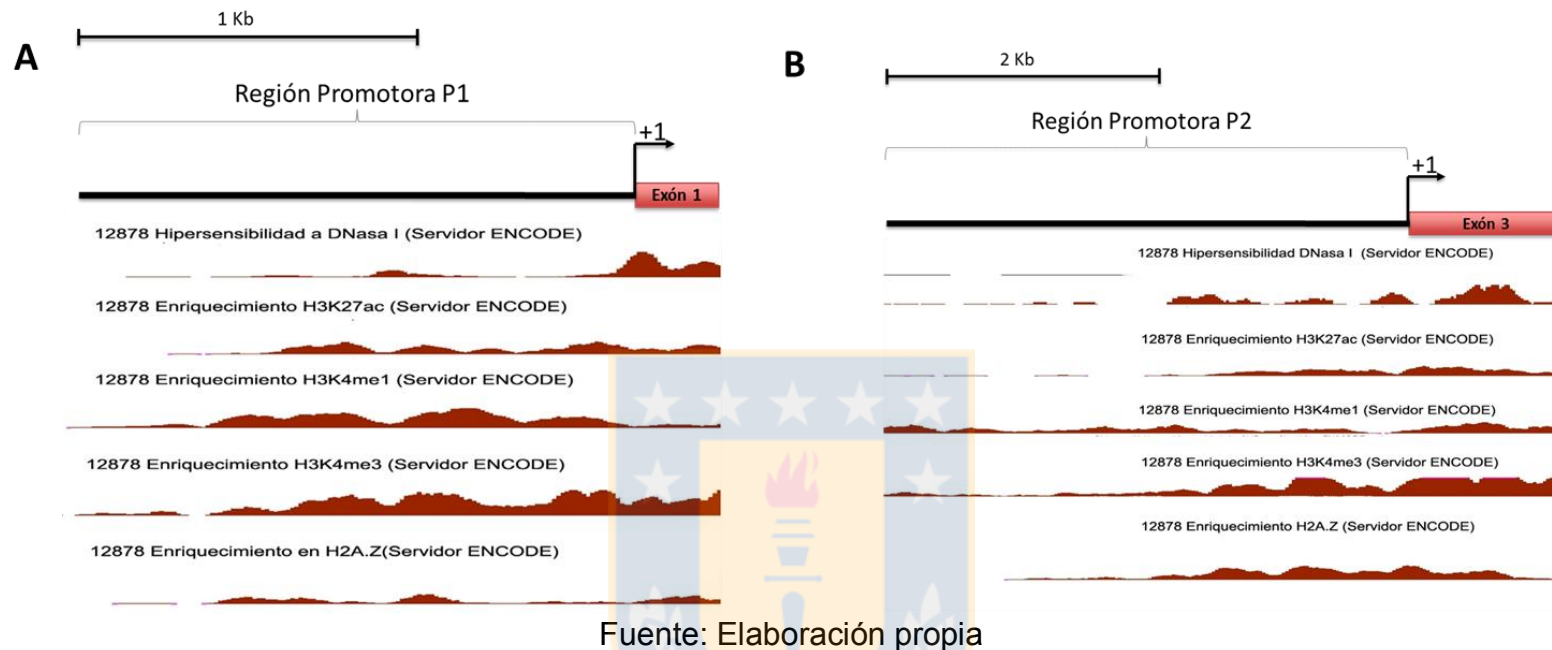


Figura 11 : El promotor P1 y P2 del gen RUNX1 presenta marcas epigenéticas asociadas a regiones Promotoras. Los datos analizados corresponden a data de ChIPseq obtenida a partir de la línea GM12878 y utilizando anticuerpos contra las modificaciones de histonas H3K27ac, H3K4me1, H3K4me3, la variante de histonas H2A.Z y ensayos de hipersensibilidad a DNasa I. La data se encuentra libremente accesible y descargable desde el servidor ENCODE. Los datos son visualizados utilizando la herramienta Genome Browser del servidor ENCODE con una ventana genómica de tamaño variable según el promotor estudiado. En ambos paneles (A y B) la región promotora se representa mediante una línea sólida negra, los exones (exón 1 y 3, respectivamente) son representados por cajas rojas, la flecha ubicada al inicio de cada exón (+1) corresponde al sitio de inicio de la transcripción. El panel A corresponde a los datos obtenidos para el promotor P1 del gen RUNX1, utilizando una ventana genómica de 1.8 Kb. El panel B corresponde al promotor P2 con una ventana genómica de 5Kb.

OBJETIVO ESPECIFICO N2: “Evaluar las marcas epigenéticas de H3K27ac, H3K4me1, H3K4me3, H2A.Z en regiones de quiebre cromosomal”.

Identificación de módulos reguladores cis en regiones de quiebre cromosomal

El intrón 5 del gen *RUNX1* ha sido identificado como la región de quiebre de este gen en la translocación 8;21. Análisis realizados por Zhang y cols 2001, determinaron que los quiebres para esta translocación, no se localizan de forma azarosa a lo largo del intrón, sino que se agrupan en 3 zonas altamente específicas denominadas BCRs1-3. Si bien para estas regiones no se han evidenciado secuencias comunes que sugieran un mecanismo de reparación homólogo, se ha determinado la presencia de elementos estructurales de la cromatina como sitios de hipersensibilidad a DNasaI y de corte a Topoisomerasa II, presentes ya sea en el interior de las zonas de quiebre o próximas a ellas (Zhang, Strissel et al. 2002), lo que sugiere que la cromatina local de intrón 5 se encuentra altamente descondensada.

Tomando en cuenta estos antecedentes, nuestro laboratorio previamente determinó la acetilación de histonas H3 y H4 a lo largo del intrón 5 del gen *RUNX1*. Stuardo y cols (Stuardo, Martinez et al. 2009), mediante ChIP comparó los niveles de acetilación de Histona H3 y H4 en líneas celulares hematopoyéticas (HL60) y epiteliales (HeLa), utilizando un anticuerpo que reconoce los residuos acetilados específicos de lisina 9 y 14 en la histona H3 y lisina 5, 8, 12, 16 y 20 de la histona H4 y posteriormente cuantificó dicho

enriquecimiento mediante PCR en tiempo real, para ello utilizó una serie de partidores (denominados por letras de la A a la U) localizados a lo largo del intrón 5 del gen *RUNX1* (**Tabla 10**). Stuardo determinó un patrón diferencial de acetilación al comparar el enriquecimiento para H3 y H4 acetiladas entre células hematopoyéticas y no hematopoyéticas, caracterizado por enriquecimiento en la acetilación de histonas H3 y H4 en distintas regiones del intrón, entre ellas zonas previamente identificadas por Zhang y cols, como regiones de quiebre cromosomal (**Figura 12**), estos antecedentes podrían sugerir la presencia de elementos *cis* reguladores en las zonas de quiebre del intrón 5.

Como primera aproximación para identificar Módulos *Cis* Reguladores (o MCR), visualizamos el conjunto de librerías desarrolladas en las etapas anteriores con la herramienta de visualización Genome Browser del servidor ENCODE (Kent, Sugnet et al. 2002) y las coordenadas genómicas que mapean el intrón 5 (**Figura 13**). Del análisis de los datos, podemos determinar que en el centro del intrón se distingue una región de aproximadamente 2,4kb, que colocaliza con la región de quiebre BCR2 descrita por Zhang y cols 2002, y que se caracteriza por presentar un peak central de Hipersensibilidad a DNasa I y enriquecimiento en la marca de histonas H3K27ac.

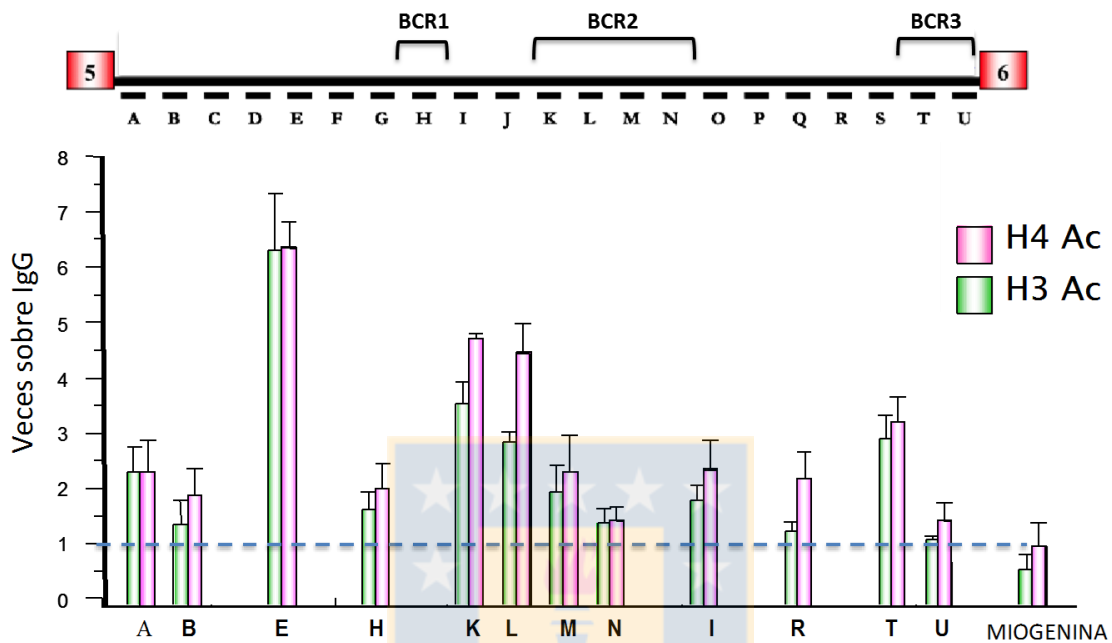
Tal como se mencionó anteriormente, ambos elementos se han mapeado frecuentemente asociados a regiones promotoras y enhancers. De hecho, se ha observado que en genes transcripcionalmente activos el sitio de inicio de la transcripción colocaliza exactamente con peaks de hipersensibilidad a DNasa I y

Tabla N° 10: Set de oligonucleótidos intrón 5 del gen *RUNX1*

Región	D	Secuencia 5'-3'	Tamaño Producto (pb)
5A (379-743)	S	AGA GTG GAC CAC AAG GGA GTA A	365
	AS	GAA CTG TAA CCC AGC CTC AGA T	
5B (1221-1638)	S	CTG TGA TCA CGG CAG CTT TTA T	418
	AS	GAA GAG CTC ACA GTC ATT CC	
5C (2350-2621)	S	GTG CAA TAG CCC CAG GTA GA	272
	AS	GGG CAT CCT GAT TTA CAT GC	
5D (3591-3893)	S	GGA ATG GGT TCT TAC CAG CA	303
	AS	TGT AGT CCC AGC TCC TCC AG	
5E (5005-5231)	S	GGG ATT ATA GGC ATC CAC CA	227
	AS	CAT ATG ACA TGA TCT CAA CCA A	
5F (6116-6400)	S	AGC AGA GCC CTG TTG CTT T	285
	AS	CCA CTT TAC CCA TCT TTC GTG	
5G (7351-7602)	S	CCA TCT CAG CTC ACT GCA AC	252
	AS	AAA TCT CAG TGG TGG CCA AG	
5H (8430-8774)	S	GAC TGA GCC TCT CCA AAA CG	345
	AS	CCA GGG CCC ATG CTA TAC TA	
5I (9956-10285)	S	TTC ATA GGT GAA AGC CAA GC	330
	AS	GGC AGA TTG ATT CCA TTG TG	
5J (11378- 11663)	S	CTG GGC ACA GAA GTG TTC CT	286
	AS	TGA AGC CAG TTC AGT GAT GC	
5K (12591-12872)	S	AGC TAG CAG GGC CAG ACA TA	282
	AS	AGC TCG CTG TCC TGT TCA TT	
5L (13447-13849)	S	TCT GAC GCT CTC TCC TGT CA	403
	AS	AAG AGG TCT CCC TCC TCA CA	
5M (14362-14687)	S	AGG GAG ACA GCC CTC ATA AAC T	326
	AS	CTC CTC TGC CTC CTC ACA GTA T	
5N (15519-15746)	S	CAG AGG GCT ACC TCT GGT CA	228
	AS	AAC TCT GCT CCC TGA GTC CA	
5O (16870- 17166)	S	GTG CCA CTC TCA ACT CAG CA	297
	AS	CAT TGT GAA ACC CCA TCT CC	
5P (18506-18720)	S	GAC TCA GCC TTG GGA TGG TA	215
	AS	TGC ACG AGG TGT CCT AAG AA	
5Q (19524-19824)	S	AAA AGG AGG CCG TCA GTC AT	301
	AS	GTG GGC AAT GGA ATT ACC TG	
5R (20389-20839)	S	CCA ACC TTG TGG TGT TGT CTT A	451
	AS	ATG GGG TTT TCT TCC TTT GA	
5S (21623- 21876)	S	GTC CCA GGG TGA AAT CTT GA	254
	AS	CCC CCG AAT GTT AAA TAG AGG	
5T (22491-22937)	S	ATA CCT CAC GTT GTG CCT GAC	447
	AS	AAT GGA ACA ACC ATT CAA AAC A	
5U (23736-23980)	S	GGG AAC AGA GGT TGT GCA AT	245
	AS	AAA CAA CCG TCA CCA GGA AG	

Fuente: Elaboración propia

Intrón 5 del gen *RUNX1*



Fuente: Elaboración propia

Figura 12: Regiones de quiebre cromosomal presentan enriquecimiento en la acetilación de Histona H3 y H4 (modificado de Stuardo y cols 2009).

Nuestro laboratorio previamente realizó la técnica de ChIP en una línea celular hematopoyética (HL60) utilizando anticuerpos específicos anti Histona H4 y H3 acetilada, que reconocen los residuos acetilados de lisina 5, 8, 12, 16 y 20 de la histona H4 y de lisina 9 y 14 en la histona H3 respectivamente. Posteriormente y para evaluar el enriquecimiento del intrón, se realizó qPCR utilizando varios set de partidores que cubren el intrón en su totalidad.

El panel superior evidencia un esquema del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Señalados con letras de la A-U se mapean de forma esquemática los set de partidores utilizados. El gráfico inferior, representa los resultados obtenidos en los ensayos de ChIP y corresponden al promedio de al menos tres experimentos independientes. Los resultados son expresados como veces sobre una inmunoprecipitación de tipo inespecífico (Veces sobre IgG). Se utilizó miogenina como control de la reacción.

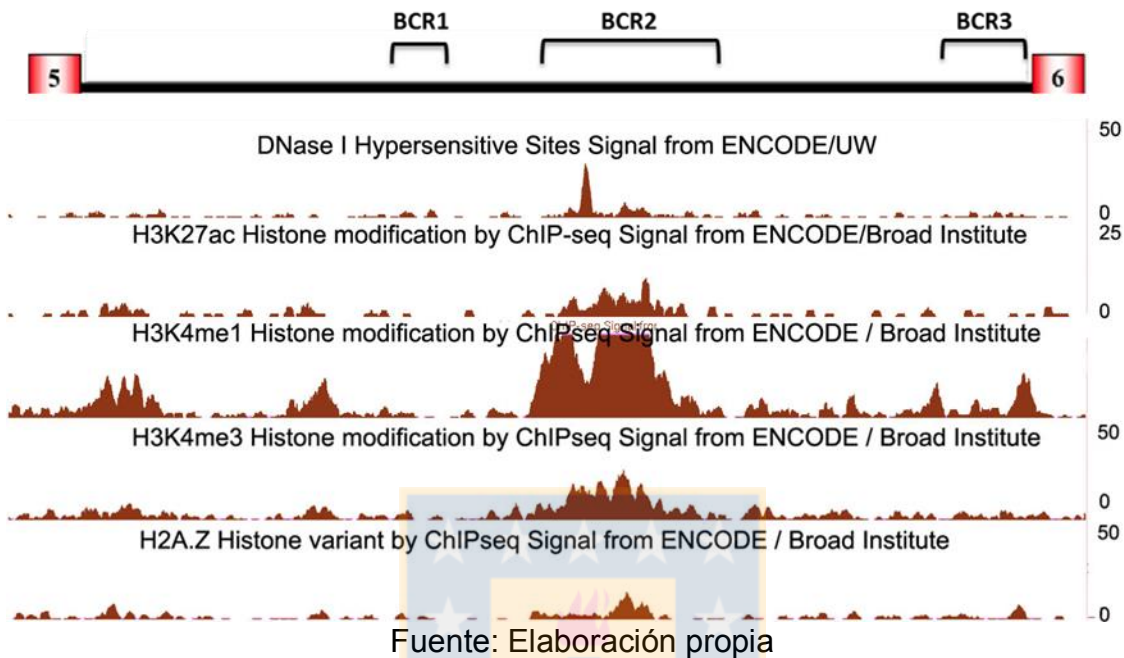


Figura 13: El intrón 5 del gen *RUNX1* presenta marcas epigenéticas asociadas a módulos *cis* reguladores.

Los datos analizados son data libremente accesible y descargable desde el servidor ENCODE. Corresponden a data de ChIPseq obtenida a partir de la línea GM12878 (células linfoblásticas) y utilizando anticuerpos contra las modificaciones de histonas H3K27ac, H3K4me1, H3K4me3, la variante de histonas H2A.Z y ensayos de Hipersensibilidad a DNasa I. Los datos son visualizados utilizando la herramienta Genome Browser del servidor ENCODE con una ventana genómica de 25 Kb (correspondientes al tamaño del intrón 5).

El panel superior se presenta una representación esquemática del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Además se resaltan con llaves negras sobre el intrón, las tres regiones de quiebre cromosomal (BCR1-3) descritas por Zhang y cols (2002). En el panel inferior se muestran los resultados obtenidos del análisis de los datos de ensayos de Hipersensibilidad a DNasa I, ChIPseq para las marcas epigenéticas de H3K27ac, H3K4me1, H3K4me3 y la variante de histonas H2A.Z. Los resultados son representados a modo de peaks y valles, donde los peaks indican las regiones altamente enriquecidas en la marca estudiada y los valles corresponden a regiones que no presentan la marca estudiada.

enriquecimiento de H3K27ac. Además de las modificaciones descritas, la región identificada presenta enriquecimiento en las modificaciones de histonas de H3K4me3, H3K4me1, y la variante de histonas H2A.Z. Todos estos elementos sugieren que, a diferencia del resto del intrón, la zona estudiada, presenta una cromatina altamente accesible y caracterizada a su vez por el enriquecimiento de marcas epigenéticas asociadas principalmente a regiones promotoras o enhancers, por este motivo denominamos a esta región como “Elemento Regulatorio Putativo” (o ERP).

OBJETIVO ESPECIFICO N3: “Identificar y caracterizar potenciales módulos *cis* regulatorios en el locus de *RUNX1*.”

ERP presenta actividad promotora

Debido a que las marcas epigenéticas analizadas pueden estar asociadas ya sea a regiones promotoras o *enhancers*, decidimos evaluar *in vitro* el rol transcripcional de la región ERP. Para cumplir este objetivo se realizaron ensayos de cuantificación de la expresión del gen reportero de luciferasa. Para ello, se amplificó una región de 2.1 kb, correspondientes a la región enriquecida, utilizando una ADN polimerasa de alta fidelidad y los partidores 5L sentido y 5K antisentido (**Tabla 10**). La región amplificada contiene el sitio de hipersensibilidad a DNasa I identificado previamente en nuestros análisis bioinformáticos y la región altamente enriquecida en marcas epigenéticas (**Figura 14**).

TCTGACGCTCTCTCCTGTACAGACGCCAACTGAGGTTCTAGGCTGGCCT
 TGCAAAGCCCTTGGACAGATGGCTGAGCAGGTCCTGGCTGGTACAGAAGG
 ACAGGGTACCCTGGCCACTCACAGCTCCTGAGTACTACTAGTACAAAAG
 CCAACCGCAGCAAGTTGTTTTGAACAGAACGTTGACCCCTCCCTCACAAAT
 AAGGCTTCCGGTTGTTTGTGTCATGACAGCCTCATTAGGCATCAGGCTC
 ACTTCTATGCACTGTTGTCCACGTGCCAGAAGCAGCCACCTGCTGAGAAG
 GGCATCTGTGAGGAGGGAGACCTTCTTAAAGTTACCTGAATGGGCAAGT
 GGCTGTTTGCCTAGACTCCACAGAGCACAGCATGGTGATGAAGATGCTG
 GTCCACTGTCTGCCAGGTAACATTGGGAGGGAAAGTTCTGTTTATTCCA
 AGACACTAGTGCCCTCCTATTACCTTGGTGACCACATTTGAGCCAGTTAG
 GAAATGAAGGAAGGGAAGGTTCCATCCTTGCAGGGATGGAGACAGGGAT
 GGAGGGCAAGGATGGACGCCAGGACAGCCTTCTGTCCGGCAACATTCCC
 CCTCCTCAATCCTTTCCGATTGAAAGGTACACAACCTCCCCCTGGCCT
 GTCTACCTTGAGTCTCTGCCTATGGCTGCCTGGCCTGGTGAGACCAGCCT
 CCCAGCTCCTCTGTGTACGCAGGCAATAGGAGAGCTGCCTTTTCAACAA
 CTGTTTTGAACCTTCTCTCCTTTTAACTCCTTCAACAGTCCCTCTCTGT
 TGTGAACACTCTGGTGTGTGAAAGCAAGACCTCAGCTCAGGGAGACAGCC
 CTCATAAAGTCCCGCAGATTACCCTGTCTTTTCCAGCCGGAGGGGGACC
 CCGCCCCACTCCCTTCCAGTCTCTTCCCCACCCCAACTTTTTTTTTTA
 ACCACTTCTGTCTCCTTCATTTTCTCAGCAGCTGCAGCGGAGAGATATAG
 TAAAGCTACTGTATAGGATTTTGAAGCAGGGCCTCCAAGAGCTTTCCAA
 ACTCCACACAAGCCAGTTATTCTCTGCCTGTTTTCTTTGGCTTCAGGTT
 AGTAACACAGAATGAGTCAAACAGTATTAACCTTCAATAGGGAAGTGGTG
 TGGTCAACCAAGCTGTTTGGCGCTAGATATGAAGATAGGTAGGGCAGG
 TAATGCGAATAGCCCATAACTGTGAGGAGGCAGAGGAGGGAGGGCGG
 CCCACTCAGAAACACTCCAACAATTAAGTCAACAGAACTTACACCATCTT
 TTTTTAGTTGGCTGAGGAATTAACACCAAGGTCTCTGACCAGCCAGCTTT
 GGGAAAGTTGGCCATAATGGGGCTTGGACAATTGAGTACTTTGTGGGTTTA
 AGGATGTCTTTTTTTTTTTTTTTTTTTTCAAATGTTGCCAAAGTCATGG
 GAACCTAGTTAGGAAGATTTCCCTACTTAGAGCATGTCACTCTGGCGT
 GGACCTTCATTCATCTCTTTCCCGGCCTGGGGTTGATCCATTCTCCCTT
 CTTCCCTTCTCCTGCCCTGCCCCACCTTCTCAGTCACATCTTTGTC
 TTCCAGTTTGGCAGGACATCCCAGGGGCCCTGGGCTTCTCTCTGAG
 GTTAGAAGATGGTGCTTTGAAAGAACAATTGCTCCTACAACCATCAAATC
 AAATTTTAGAAGCTGATAAAGACCTAAAAATCATTCAATCTCCTCTTC
 TTTCTCCTTCTGTAGATGAGGAATCTGAGGGTCTAAGAGGAGACATGAGC
 TGCTGAGATTGACTGAGCTAGCAGGGCCAGACATAGACTCCACGGGCAGG
 CCCGCACTCTTCCACCAGGTCAGGCCACTCCTCGCCAGGGCTCAAGA
 TTCCTTTTAGAGTTCAACAGAGCAGCTCCCGAGAACAGACCAGCCCTCAA
 GACTCAAATTTCAAGTGGTTTCCAGTCTTTTCCCTGCCTCAGCTACCT
 ATTTATTTTGTAGACTGGCTGAATCACCCGAAGAATAGAACGGGTTTCTCT
 GAAGCTGCTCCCGCAGCATGTCAGCAAATGAACAGGACAGCGAGCT

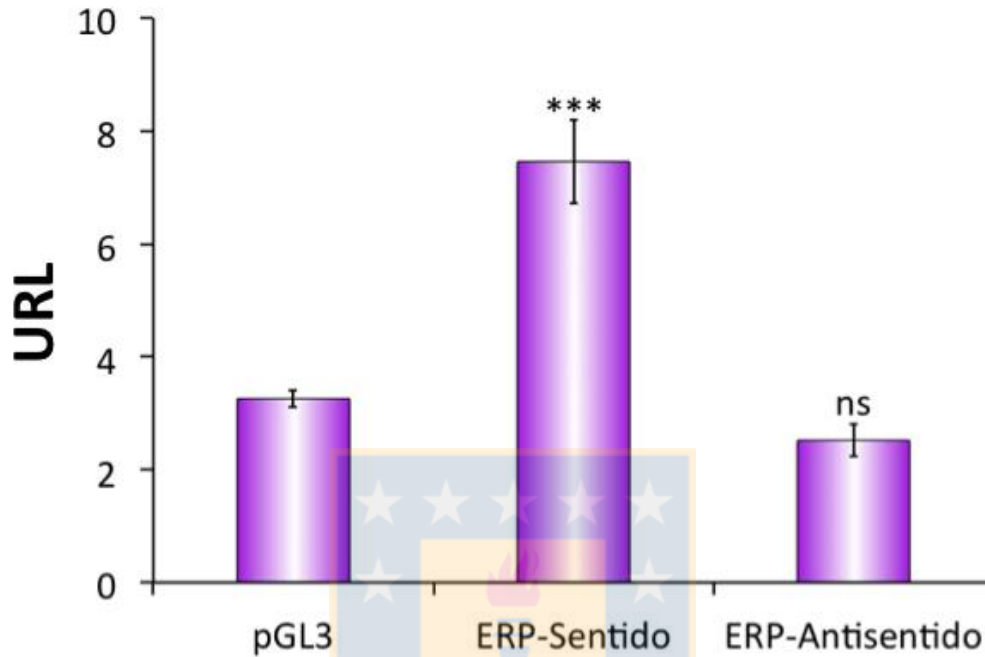
Fuente: Elaboración propia

Figura 14: Secuencia y características de la región clonada.

Utilizando el visor Genome Browser se delimitó una región de 2.1kb que presentó enriquecimiento en las marcas epigenéticas de H3K27ac, H3K4me1, H3K4me3, la variante de histonas H2A.Z e hipersensibilidad a DNasa I. Posteriormente y utilizando sus coordenadas genómicas de inicio y termino se descargó su secuencia de ADN desde el servidor ENCODE. Destacado en color amarillo, se resalta la región de hipersensibilidad identificada mediante nuestros análisis bioinformáticos. Además, en color celeste, se muestra que región presentó el mayor peak de hipersensibilidad.

Posteriormente, la región ERP fue clonada en el vector reportero pGL3-Basic que presenta la particularidad de no poseer promotor para la expresión del gen reportero de luciferasa. Tomando en cuenta que las regiones promotoras modulan la expresión génica de modo orientación dependiente mientras que los *enhancers* lo hacen de forma independiente de distancia y sentido, se construyeron dos vectores distintos, generados mediante el clonamiento de la región ERP en orientación sentido y antisentido corriente arriba del marco de lectura del gen reportero. De esta manera, al evaluar la expresión del gen reportero en el modelo celular HeLa, mediante transfecciones transientes dependientes de la actividad transcripcional de la región ERP, podemos identificar si se trata de un elemento promotor o *enhancer*.

Nuestros resultados demuestran que al utilizar el vector que contiene la región ERP en orientación antisentido no se obtuvieron cambios significativos en la expresión del gen reportero de luciferasa en comparación con la actividad observada para el vector pGL3 Basic vacío (**Figura 15**). Sin embargo, al evaluar la región ERP en orientación sentido, se observa un aumento significativo en la expresión del gen reportero ($p < 0.005$, $n = 12$). Estos resultados dan cuenta que por una parte, la región ERP modula la expresión del gen reportero de luciferasa y que esta modulación además es dependiente de la orientación de la región, por lo cual concluimos que la región ERP presenta actividad promotora *in vitro*.



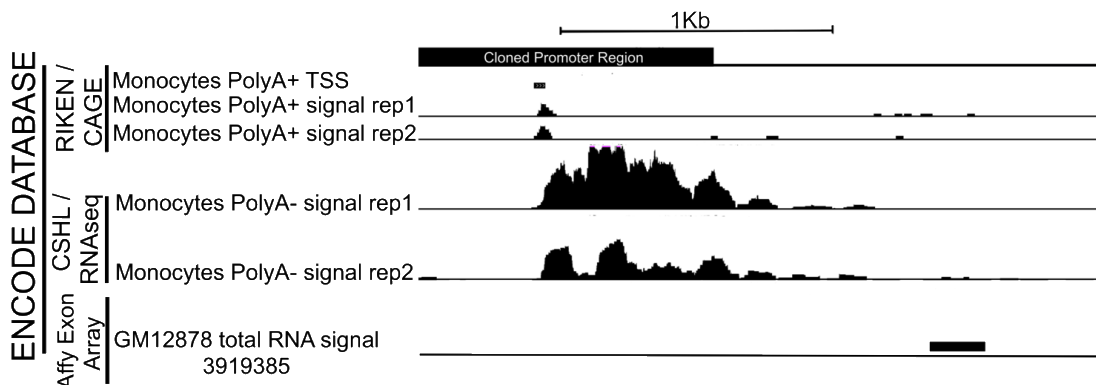
Fuente: Elaboración propia

Figura 15: La región ERP modula la expresión del gen reportero de luciferasa de manera orientación dependiente.

Células HeLa fueron transfectadas utilizando 200 ng de los plásmidos pGL3 basic, pGL3 con la región ERP en orientación sentido o antisentido con respecto al gen reportero de luciferasa. Todas las transfecciones fueron normalizadas con la cotransfección del plásmido de expresión de renilla y los datos son expresados como unidades relativa de luciferasa (URL). La actividad reportera fue cuantificada 18 horas post-transfección. Cada barra representa el promedio más el error estándar de la media (e.e.m) de cuatro experimentos independientes, cada uno realizado por triplicado. Los asteriscos representan la significancia estadística determinada por el *t-test* (***) $p < 0.001$, ns= no significativo).

Identificación in silico e in vivo de un potencial ARN expresado desde el promotor ERP

Un elemento promotor es el lugar físico donde son reclutados factores generales y específicos de la transcripción para el inicio de la expresión génica. Tomando en cuenta que la región ERP *in vitro* presenta actividad promotora, decidimos investigar si un transcrito es expresado a partir de ella. Primeramente tomamos las coordenadas genómicas de ERP y a partir de ellas se evaluaron los perfiles de expresión en distintas líneas celulares, para lo que se utilizó data libremente accesible descargadas desde el servidor ENCODE en las bases de datos RIKEN/CAGE, CSHL y Affy Exon Array. Nuestros análisis detectaron la expresión de un pequeño ARN polyA+ de aproximadamente 200 pb, sintetizado desde el centro de la región ERP y, sorprendentemente, a partir de la hebra positiva (*RUNX1* es transcrito desde la hebra negativa). Este fragmento de ARN fue posteriormente utilizado como anzuelo en el servidor EST Data Base de PubMed para identificar secuencias de EST que presentaran un 100% de homología de secuencia con este anzuelo. De este modo, se detectaron 9 secuencias de EST, expresadas desde distintos tipos de tejidos (la mayoría neoplásicos) y se ensambló mediante el solapamiento de estos, un transcrito putativo 776 pb que fue denominado de forma arbitraria NAT-RUNX1 (Natural Antisense RUNX1) (**Figura 16**). Posteriormente se sintetizaron partidores dirigidos contra NAT-RUNX1 y se realizó RT-PCR en células hematopoyéticas HL60 (leucemia mieloide) y Jurkat (Leucemia Linfoide), utilizando como



GeneBank	Organ	Tissue type	pb
AA984901.1	spleen	spleen	313
AI810863.1	prostate	prostate	543
AI798118.1		B-cell, chronic lymphocytic leukemia	542
AI863417.1	brain	different tumor types	534
BX111268.1	brain	different tumor types	465
AW204843.1	colon	colon	377
BU684610.1	Lung	Primary Lung Cystic Fibrosis Epithelial Cells	772
BU689215.1	Lung	Lung	708
AW516704.1	ovary	papillary serous ovarian metastasis	360
bait	GM12878	exon array signal 3919385	206

Fuente: Elaboración propia

Figura 16 : Identificación *in silico* de un ARN putativo expresado desde la región ERP.

Utilizando las coordenadas genómicas de la Región ERP, con una ventana genómica de 2.5 Kb, se analizó data descargada desde el servidor ENCODE en las librerías RIKEN CAGE, CSHL y Affy Exon Array. Se identificó una secuencia corta de ARN polyA+, que fue posteriormente utilizada como anzuelo en la base de datos de Expressed Sequence Tags (ESTs) del servidor PubMed.

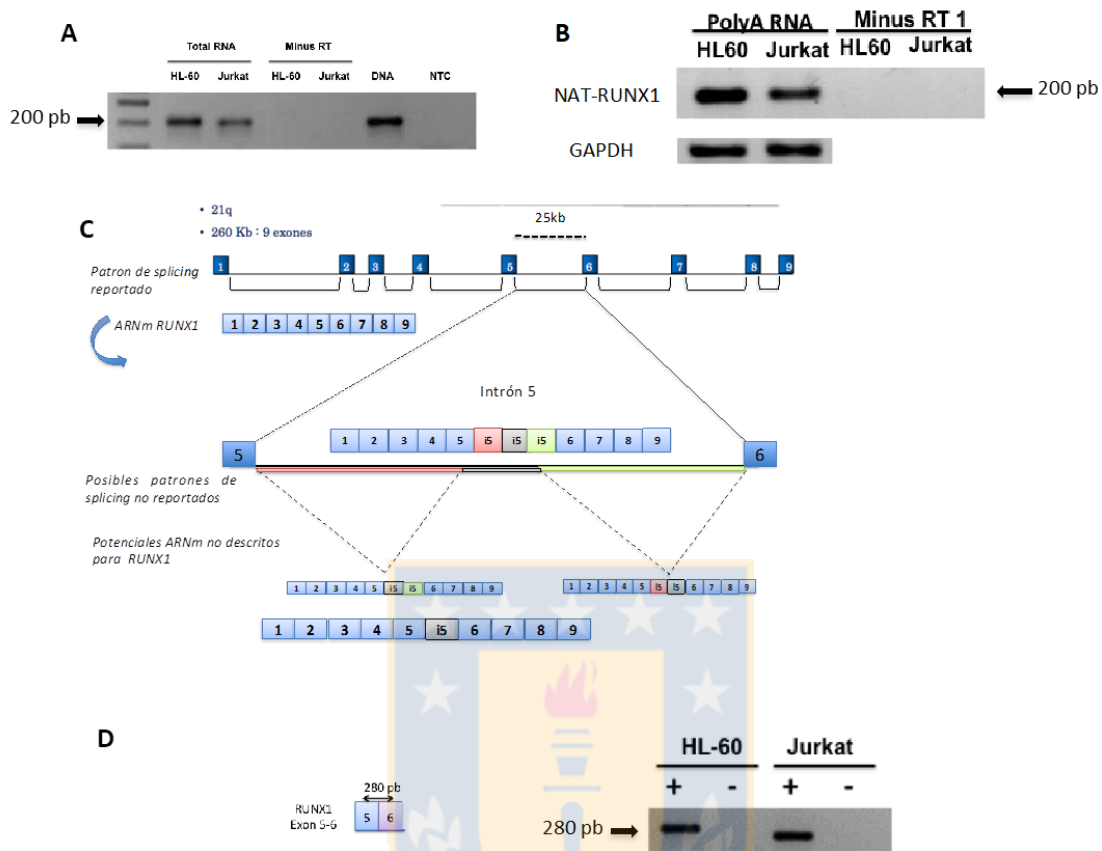
En el panel superior se representa un esquema de la región ERP (caja de color negro), se señalan también los resultados obtenidos para la ubicación de los distintos ARNs identificados en las bases de datos de RIKEN, CSHL y Affy Exón Array. En la tabla del panel inferior se detallan los EST (junto a su identificador GeneBank) obtenidos al utilizar como anzuelo una región presente en todos los ESTs identificados en las librerías RIKEN CAGE, CSHL y Affy Exon Array.

templado ARN total. Nuestros resultados evidencian que ambas líneas celulares expresan NAT-RUNX1, descartándose además que los resultados obtenidos sean productos de la amplificación de ADN contaminante, ya que en ambos controles de pureza de la reacción, no se produjo amplificación (**Figura 17 panel A**). Debido a estos resultados, se decidió evaluar si el potencial transcrito detectado corresponde a un transcrito distinto a RUNX1 o es una amplificación del ARN de *RUNX1* no procesado por la maquinaria de splicing.

Para ello, realizamos un aislamiento de ARNs polyA+. Nuestros resultados evidencian que es posible detectar la expresión de NAT-RUNX1 en ambas líneas celulares y que por ende NAT-RUNX1 corresponde a un transcrito polyA+ (**Figura 17 panel B**). Finalmente, para descartar que NAT-RUNX1 sea generado como splicing alternativo entre el exón 5 y 6 de RUNX1, se utilizó un set de partidores que se localizan en dichos intrones y cuyo producto de amplificación corresponde a un amplicon de 280 pares de bases. Nuestros resultados demuestran que al realizar RT-PCR con ARN polyA+ en ambas líneas celulares se obtiene un producto único de amplificación de 280 pb, lo que nos permite sugerir que NAT-RUNX1 corresponde a un transcrito que es expresado de forma independiente a *RUNX1* (**Figura 17 panel D**).

La proteína RUNX1 se asocia al elemento regulador ERP en regiones de quiebre cromosomal

La estructura de la cromatina y su funcionalidad pueden ser afectadas por múltiples modificaciones de tipo post traduccional en el extremo amino terminal



Fuente: Elaboración propia

Figura 17: NAT RUNX1 es expresado en células hematopoyéticas.

Se sintetizaron oligonucleótidos dirigidos contra el mensajero putativo ensamblado a partir de los análisis de EST (denominado NAT-RUNX1). Posteriormente se realizó RT-PCR utilizando como templado ARN total (panel A), ARN poly A+ (panel B y D) extraído desde células HL60 (promielocitos) y Jurkat (Linfoblastos). Para todas las reacciones se realizó un control de pureza de la reacción para descartar contaminación con ADN genómico, este control contiene el templado utilizado para cada síntesis de cDNA y todos los reactivos utilizados excepto la enzima transcriptasa reversa (Minus RT). Se sintetizó un set de partidores que amplifica de forma específica NAT-RUNX1 generando un amplicon de 200 pb. Para la reacción con RNA total (Panel A). se utilizó también un control de amplificación positivo utilizando como templado ADN genómico. El panel C contiene una representación esquemática del *splicing* canónico de RUNX1 y de las posibles isoformas de *splicing* alternativo en caso de retener parte o la totalidad del intrón 5, los exones se representan por cuadrados azules (1-9), intrones por la línea continua de color azul entre los exones.

de las histonas, donde la acetilación de los residuos de lisina ha sido una de las más ampliamente estudiadas, por su asociación con activación transcripcional (Kimura 2013, Rajagopal, Ernst et al. 2014, Pradeepa, Grimes et al. 2016). La acetilación de histonas es llevada a cabo por una familia de enzimas denominadas Histonas Acetiltransferasas (HATs) (Kitabayashi, Yokoyama et al. 1998). Para efectuar su actividad catalítica, las HATs son reclutadas por factores de transcripción tejido específicos a las regiones reguladoras de sus genes blancos, acetilando los residuos de lisina de las histonas presentes, orquestando de este modo una estructura de cromatina abierta y accesible a la maquinaria transcripcional (Simon, North et al. 2011). Previamente nuestro laboratorio comparó los niveles de acetilación de Histona H3 y H4 en líneas celulares hematopoyéticas y epiteliales, utilizando un anticuerpo que reconoce los residuos acetilados específicos de lisina 9 y 14 en la histona H3 y lisina 5, 8, 12, 16 y 20 de la histona H4 (Stuardo et al., 2009) (**Figura 12**). Los resultados muestran la presencia de un patrón diferencial de acetilación en células HL60. Dicho patrón se caracteriza por hiperacetilación de la histona H3 y H4 en regiones descritas como sitios de quiebre por Zhang y cols (Zhang, Strissel et al. 2002).

En adición a estos resultados, nuestros análisis identificaron la presencia de un sitio de hipersensibilidad a DNasa I colocalizando con una de las regiones hiperacetiladas, lo que permite sugerir que en células hematopoyéticas, la estructura local de cromatina presente en esta región de quiebre es altamente accesible. Tomando en cuenta que identificamos un potencial elemento regulador con actividad promotora *in vitro* (región ERP) resulta importante determinar el

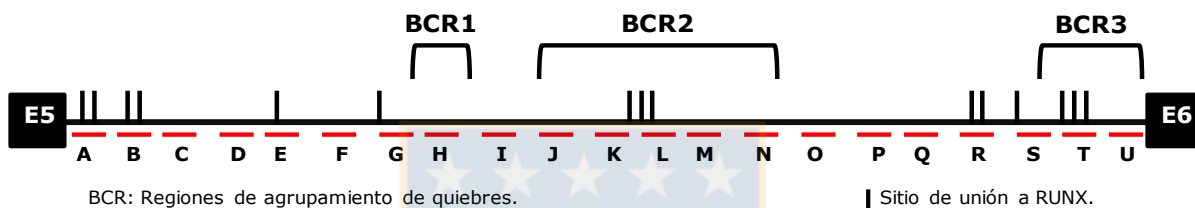
control transcripcional de dicho promotor, por ello decidimos evaluar qué factores regulan la estructura de cromatina presente en la región ERP.

Primeramente realizamos la búsqueda *in silico* de sitios de unión a factores de transcripción a lo largo del intrón 5 y con especial énfasis en la región ERP. Para ello se utilizó el servidor TF Search y se establecieron dos criterios de búsqueda: que los sitios de unión presenten al menos un 92% de homología a su secuencia consenso y que correspondan a motivos de unión a factores de transcripción hematopoyéticos. Los resultados obtenidos demuestran que utilizando estos criterios sólo se identificaron sitios de unión a proteínas RUNX presentes en el intrón 5. Estos motivos de unión se localizan a modo de clusters en regiones previamente determinadas como hiperacetiladas y en el elemento regulador ERP (**Figura 18**).

Considerando estos resultados, decidimos determinar la asociación *in vivo* de RUNX1 al intrón 5 del gen *RUNX1* mediante la técnica de ChIP en células HL60, la detección se realizó mediante qPCR. Tomando en cuenta que los sitios de unión a RUNX identificados, pueden ser amplificados utilizando los partidores previamente empleados por Stuardo y cols 2009 se emplearon los mismos oligonucleótidos para la detección de las regiones enriquecidas (**Tabla 10**).

Al analizar nuestros resultados podemos observar que en células HL60, de todas las regiones estudiadas, la proteína RUNX1 presenta unión a dos regiones, la región J de forma leve y en contraste presenta una alta asociación a la región L (**Figura 19**).

La región L, se localiza en el interior de la región ERP clonada y muy próxima a la región hipersensible a DNasa I. El factor de transcripción RUNX1 es



Fuente: Elaboración propia

Figura N°18: Distribución de sitios putativos de unión a RUNX identificados *in silico*.

Se realizó un análisis bioinformático utilizando el servidor TF Search donde se establecieron 2 criterios de búsqueda: a) Motivos de unión a factores de transcripción con al menos un 92% de consenso de homología de secuencia y b) que correspondieran a sitios de unión a factores de transcripción hematopoyéticos. Bajo estos criterios, sólo se identificaron 13 sitios de unión a proteínas RUNX. En la figura, en líneas negras verticales, se esquematiza la ubicación de los motivos consenso a RUNX con respecto a los oligonucleótidos sintetizados (nombrados de la A a la letra U) y se marcan, con llaves negras, los sitios de quiebre para la t(8;21) descritos por Zhang y cols 2002 (BCR1-3).

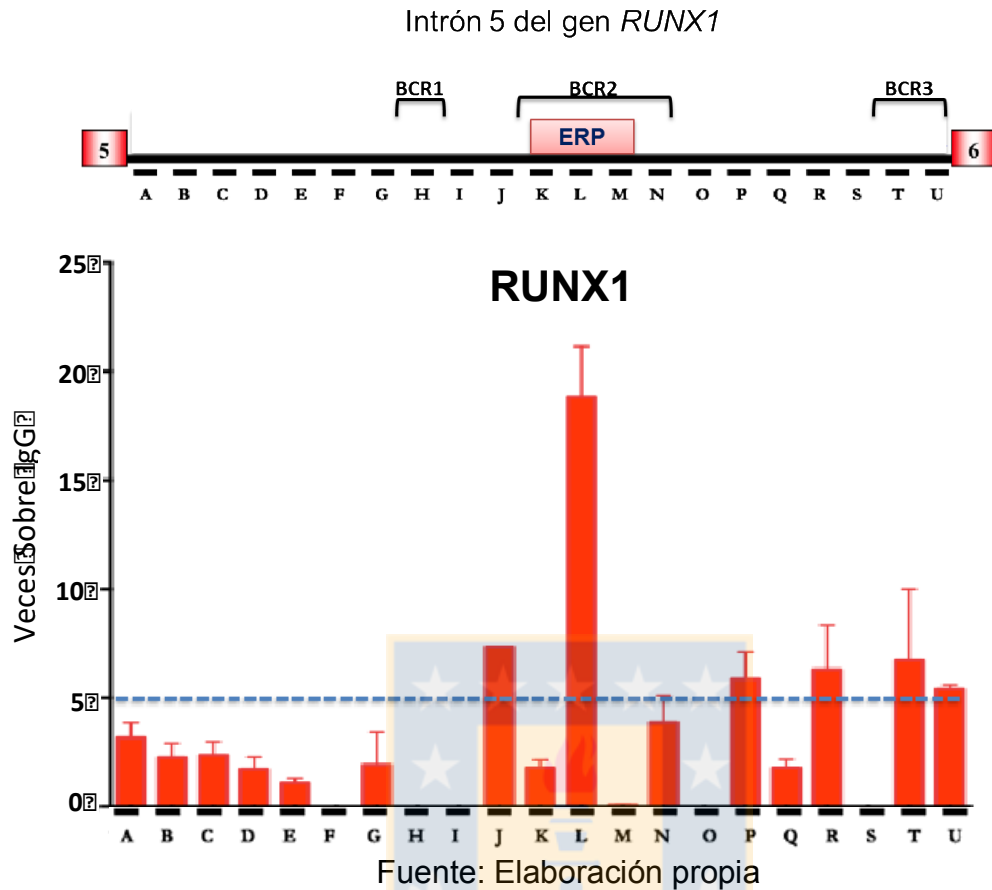


Figura 19: La proteína RUNX1 se asocia a la región ERP en el intrón 5 del gen *RUNX1*.

Se realizó inmunoprecipitación de cromatina en células hematopoyéticas HL60 utilizando un anticuerpo específico anti-RUNX1 y un anti-IgG como control de especificidad de la reacción. El ADN de los complejos inmunoprecipitados fue aislado mediante una extracción fenólica. Posteriormente se realizó PCR en tiempo real. El panel superior representa un esquema del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Señalados con letras de la A-U ilustran de forma esquemática los set de partidores utilizados. Además destacan con llaves negras, los sitios de quiebre para la t(8;21) descritos por Zhang y cols 2002 (BCR1-3) y en un rectángulo de color rojo el elemento regulador putativo (ERP) identificado previamente. En el panel inferior, se representan los resultados obtenidos de la Inmunoprecipitación de cromatina. Estos fueron graficados de forma arbitraria, utilizando como parámetro de comparación “las veces sobre la IgG”, que corresponde al cociente entre la cuantificación de la inmunoprecipitación de RUNX1/IgG. Se determinó como umbral de especificidad cinco veces sobre el valor de la IgG (línea discontinua color celeste) Los resultados presentados corresponden al promedio de al menos tres experimentos realizados de forma independientes con triplicados técnicos para cada determinación (n>9).

un importante regulador de la estructura de cromatina el cual ha sido propuesto como factor organizador de la cromatina gracias a su capacidad de reclutar a una serie de enzimas, entre las que destacan las HATs (Cohen, 2009; Ito, 2004; Kitabayashi et al., 1998a; Mikhail et al., 2006; Oakford et al., 2010; Stuardo et al., 2009).

Las HATS pCAF, CBP y p300 se localizan en regiones de quiebre cromosomal y en la región ERP

Nuestro laboratorio previamente reportó un perfil diferencial de acetilación caracterizado por enriquecimiento de las histonas H3 y H4 en regiones acotadas del intrón. Al analizar la data obtenida por Stuardo y cols (Stuardo, Martinez et al. 2009), podemos determinar que la zona donde se localiza el elemento regulador ERP presenta acetilación de histona H3 y H4 en células hematopoyéticas. La acetilación de histonas es catalizada por enzimas de tipo Histonas Acetiltransferasas donde pCAF, p300 y CBP han sido ampliamente caracterizadas acetilando los residuos de lisina de las histonas H3 y H4 previamente evaluado por Stuardo y cols. Tomando en consideración estos antecedentes, decidimos evaluar la asociación de estas Histonas Acetiltransferasas al intrón 5 del gen *RUNX1*. Para ello realizamos análisis de CHIP y evaluamos su asociación mediante PCR en tiempo real.

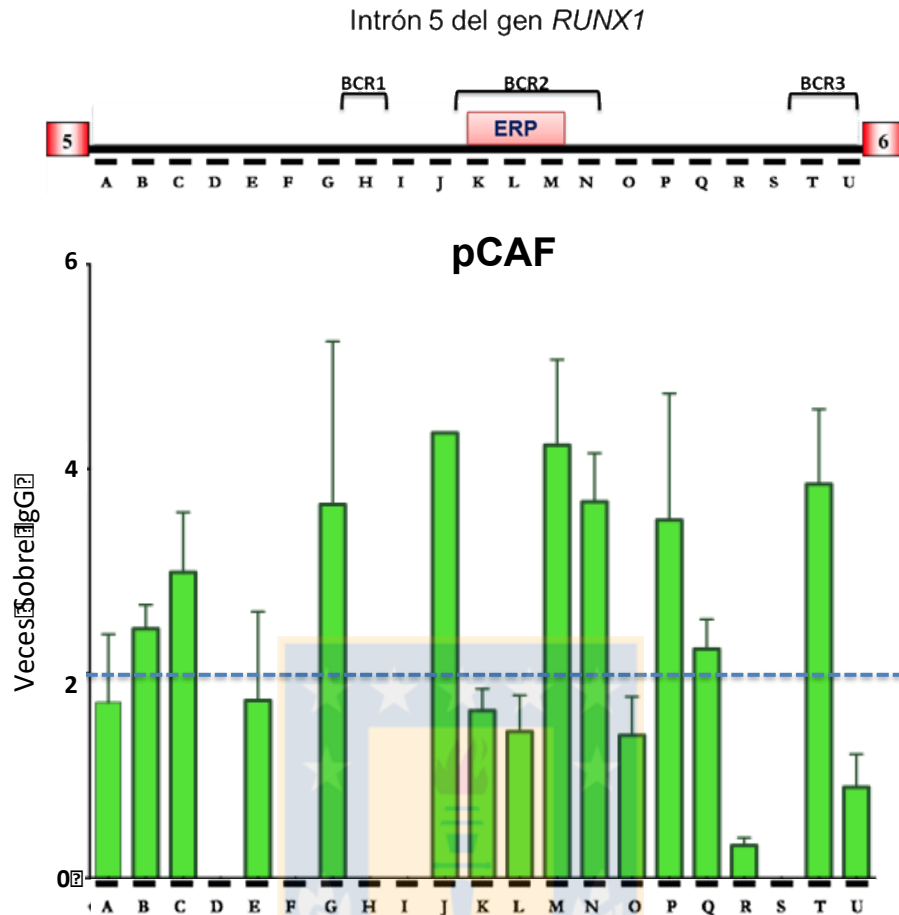
Nuestros resultados evidencian que existe un patrón diferencial de unión para las HATs evaluadas. De hecho, al determinar la asociación de pCAF al intrón 5, podemos distinguir que aparentemente el extremo 3' de este intrón, presenta asociación a pCAF, destacando principalmente las regiones M, N y T y en menor

medida a las regiones G, J y P (**Figura 20**). Todas estas regiones se localizan al interior de regiones de quiebre o próximas a estas, e interesantemente la región 5M corresponde al segmento del elemento ERP que presenta mayor hipersensibilidad a DNasa I.

Al evaluar la asociación de CBP podemos observar que solo dos de las regiones mapeadas presentan enriquecimiento en CBP. Estas regiones corresponden a las regiones N y P. Es importante destacar que, a pesar de que la región N no forma parte de ERP, de igual modo, se localiza adyacente a ella y está presente en el interior de una región de quiebre (**Figura 21**). Cuando se evaluó la asociación de p300 al intrón 5, nuestros resultados evidenciaron un enriquecimiento en las regiones B, C, L, entre estas regiones destaca la región L, ya que forma parte de la región reguladora ERP y además presenta asociación de RUNX1 (**Figura 22**).

En la región ERP, RUNX1 y p300 forman parte del mismo complejo

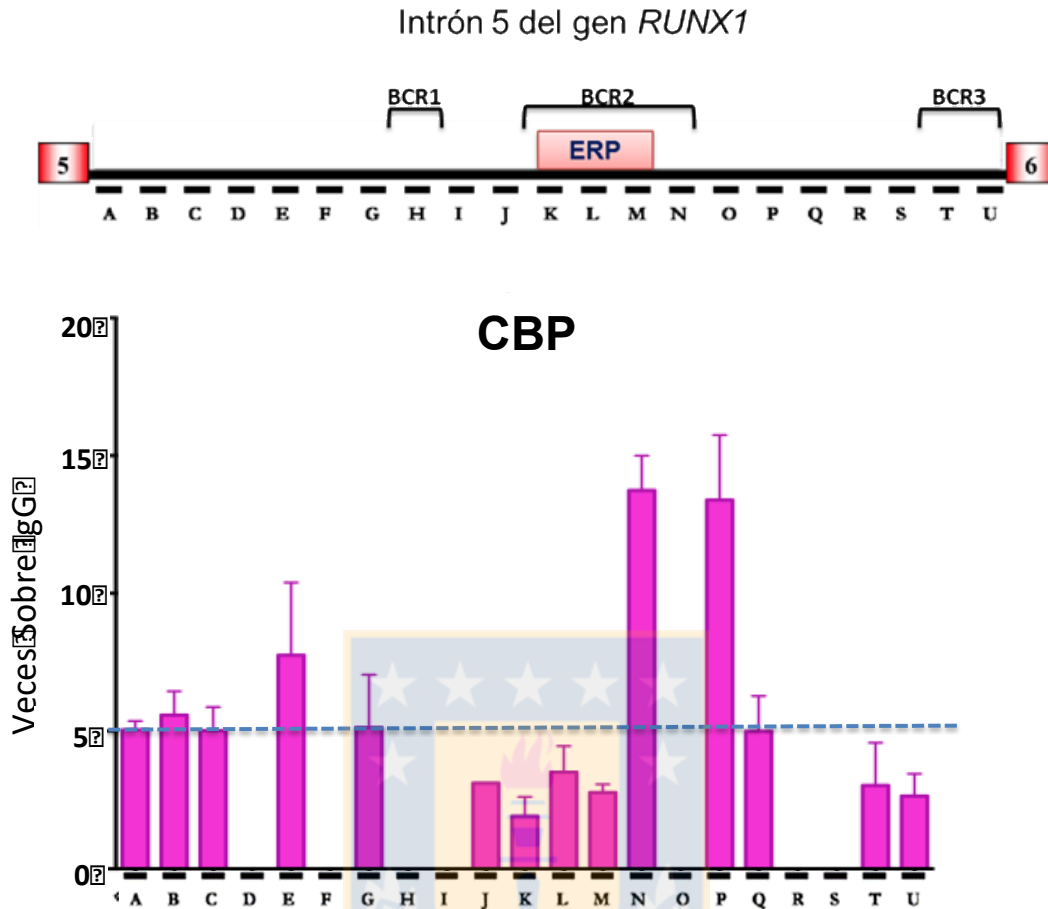
Previamente determinamos que tanto RUNX1 como p300 se asocian a la región ERP (**Figura 19 y 22**). A la fecha, varios estudios han determinado la asociación entre el factor de transcripción RUNX1 y p300 a regiones promotoras y *enhancer* de genes transcripcionalmente activos (Oakford et al., 2010; Lichtinger et al., 2010; Kitabayashi et al., 1998b), lo que sugeriría que RUNX1 puede actuar como factor organizador de la estructura de cromatina al reclutar a p300 y, eventualmente, a otras HATs a regiones *cis* reguladoras, para de este modo favorecer la adquisición de una estructura de cromatina accesible a la maquinaria enzimática (Chen et al., 2001; Lichtinger et al., 2010). Debido a estos



Fuente: Elaboración propia

Figura 20: pCAF se asocia a regiones acotadas del intrón 5 del gen *RUNX1*.

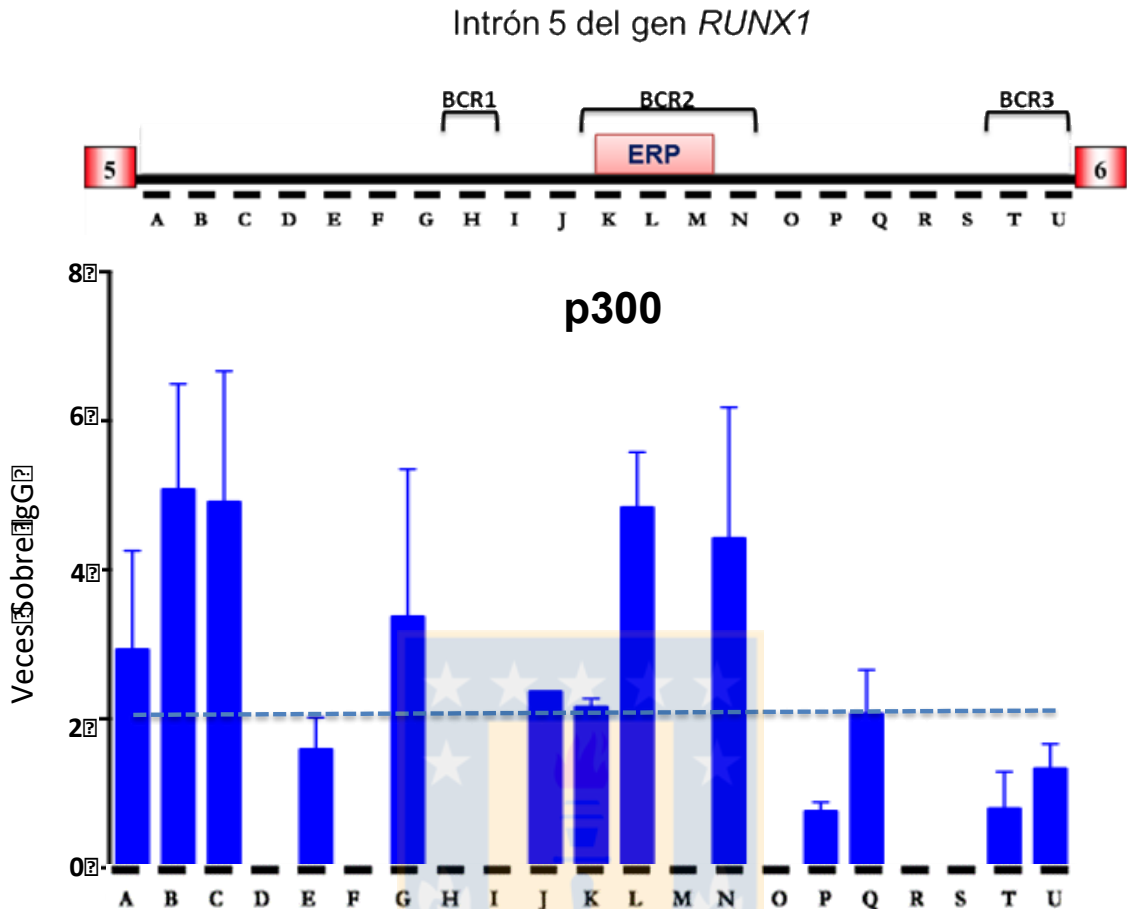
Se realizó inmunoprecipitación de cromatina en células hematopoyéticas HL60 utilizando un anticuerpo específico anti-pCAF y anti-IgG como control de inespecificidad de la reacción. El ADN de los complejos inmunoprecipitados fue aislado mediante una extracción fenólica. Posteriormente se realizó PCR en tiempo real. El panel superior representa un esquema del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Señalados con letras de la A-U ilustran de forma esquemática los set de partidores utilizados. Además destacan con llaves negras, los sitios de quiebre para la t(8;21) descritos por Zhang y cols 2002 (BCR1-3) y en un rectángulo de color rojo el elemento regulador putativo (ERP) identificado previamente. En el panel inferior, se representan los resultados obtenidos de la Inmunoprecipitación de cromatina. Estos fueron graficados de forma arbitraria, utilizando como parámetro de comparación “las veces sobre la IgG”, que corresponde al cociente entre la cuantificación de la inmunoprecipitación de pCAF/IgG. Se determinó como umbral de especificidad dos veces sobre el valor de la IgG (línea discontinua celeste). Los resultados presentados corresponden al promedio de al menos tres experimentos realizados de forma independientes con triplicados técnicos para cada determinación (n>9).



Fuente: Elaboración propia

Figura 21: CBP se asocia a regiones acotadas del intrón 5 del gen *RUNX1*.

Se realizó inmunoprecipitación de cromatina en células hematopoyéticas HL60 utilizando un anticuerpo específico anti-CBP y anti-IgG como control de inespecificidad de la reacción. El ADN de los complejos inmunoprecipitados fue aislado mediante extracción fenólica. Posteriormente se realizó PCR en tiempo real. El panel superior representa un esquema del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Señalados con letras de la A-U ilustran de forma esquemática los set de partidores utilizados. Además destacan con llaves negras, los sitios de quiebre para la t(8;21) descritos por Zhang y cols 2002 (BCR1-3) y en un rectángulo de color rojo el elemento regulador putativo (ERP) identificado previamente. Los resultados obtenidos (panel inferior) fueron graficados de forma arbitraria, utilizando como parámetro de comparación “las veces sobre la IgG”, que corresponde al cociente entre la cuantificación de la inmunoprecipitación de CBP/IgG. Se determinó como umbral de especificidad cinco veces sobre el valor de la IgG. Los resultados presentados corresponden al promedio de al menos tres experimentos realizados de forma independientes con triplicados técnicos para cada determinación (n>9).



Fuente: Elaboración propia

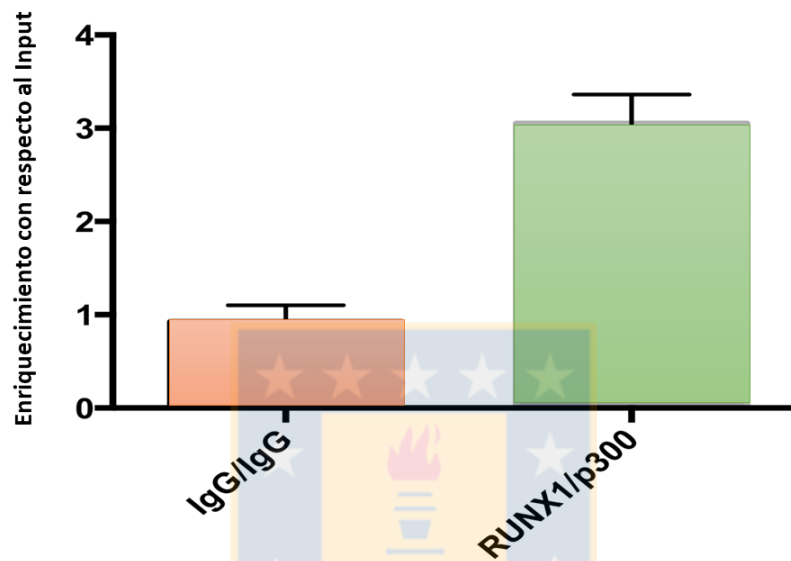
Figura 22: p300 se asocia a la región ERP en el intrón 5 del gen *RUNX1*.

Se realizó inmunoprecipitación de cromatina en células hematopoyéticas HL60 utilizando un anticuerpo específico anti-p300 y anti-IgG como control de inespecificidad de la reacción. El ADN de los complejos inmunoprecipitados fue aislado mediante una extracción fenólica. Posteriormente se realizó PCR en tiempo real. El panel superior representa un esquema del intrón 5, donde en cajas de color rojo se representan los exones 5 y 6. La línea horizontal continua negra representa al intrón 5. Señalados con letras de la A-U se ilustran de forma esquemática los set de partidores utilizados. Además destacan con llaves negras, los sitios de quiebre para la t(8;21) descritos por Zhang y cols 2002 (BCR1-3) y en un rectángulo de color rojo el elemento regulador putativo (ERP) identificado previamente. Los resultados obtenidos (panel inferior) fueron graficados de forma arbitraria, utilizando como parámetro de comparación “las veces sobre la IgG”, que corresponde al cociente entre la cuantificación de la inmunoprecipitación de p300/IgG. Se determinó como umbral de especificidad dos veces sobre el valor de la IgG. Los resultados presentados corresponden al promedio de al menos tres experimentos realizados de forma independientes con triplicados técnicos para cada determinación (n>9).

antecedentes, decidimos evaluar si efectivamente, RUNX1 y p300 forman parte del mismo complejo en la región ERP. Para ello, en células HL60, se empleó la técnica de Re-Inmuno Precipitación de Cromatina o ReChIP donde se utilizó anticuerpos específicos anti p300 y RUNX1. Nuestros resultados muestran que al analizar la región L, encontramos enriquecimiento en la asociación de RUNX1/p300, lo que nos permite concluir que efectivamente, en la región ERP existe formación y asociación de un complejo multiproteico, compuesto tanto por RUNX1 como por p300 (**Figura 23**).

RUNX1 modula la actividad transcripcional de la región ERP.

Debido a que previamente se determinó que RUNX1 se une *in vivo* a la región ERP y forma complejos con la histona acetiltransferasa p300, decidimos evaluar el potencial rol regulador del factor de transcripción RUNX1 sobre la actividad transcripcional de la región ERP. Para cumplir este objetivo, se utilizaron células HeLa, que se caracterizan por no presentar expresión de la proteína RUNX1, y realizamos cotransfecciones con el plásmido que contiene el elemento ERP en orientación sentido con respecto al gen reportero (que presentó actividad transcripcional), y concentraciones crecientes de un plásmido de expresión para RUNX1. Al utilizar concentraciones del orden de 25 a 50 ng de vector de expresión para RUNX1 se observó una disminución en la expresión del gen reportero de luciferasa altamente significativa ($p < 0,05$ y $p < 0,5$ respectivamente, $n=12$). Interesantemente, al utilizar concentraciones de 100 y 200 ng no se observaron cambios significativos al compararlos con la región ERP sentido embargo al utilizar 500 ng totales del vector de expresión



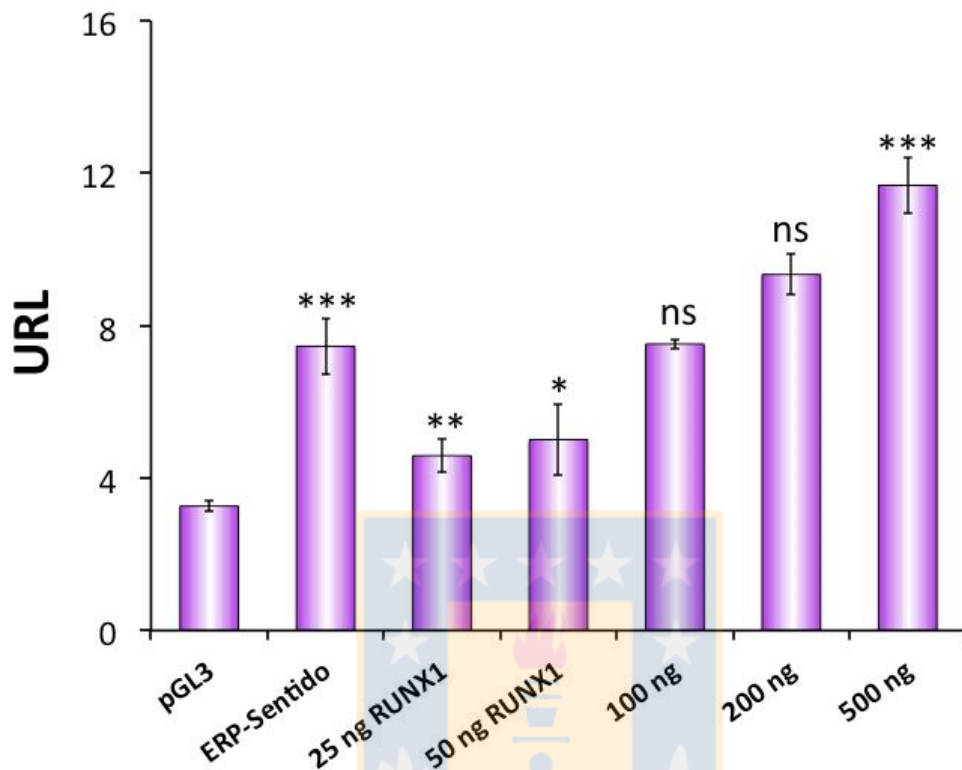
Fuente: Elaboración propia

Figura 23: La proteína RUNX1 y p300 forman parte del mismo complejo en el promotor ERP.

Se realizó Re-inmunoprecipitación de cromatina (ReChIP) en células hematopoyéticas HL60 utilizando anticuerpos específicos anti-RUNX1, anti-p300 y anti-IgG como control de inespecificidad de la reacción. Utilizando como base el primer inmunoprecipitado realizado con el anticuerpo antiRUNX1 y un control con IgG, posteriormente se realizó la segunda Inmunoprecipitación con anticuerpos anti-p300 y Anti-IgG. El ADN de los complejos inmunoprecipitados fue aislado mediante una extracción fenólica. Posteriormente se realizó PCR en tiempo real utilizando partidores que amplifican la región 5L. Los resultados obtenidos fueron graficados utilizando como parámetro de comparación “el enriquecimiento con respecto al input”. Se determinó como umbral de especificidad una vez sobre el valor de la IgG. Los resultados presentados corresponden al promedio de al menos tres experimentos realizados de forma independientes con triplicados técnicos para cada determinación (n=9).

observamos un aumento en la expresión del gen reportero de luciferasa altamente significativo ($p < 0.005$ $n = 12$) al comparar con la actividad del plásmido que contiene la región ERP Sentido (**Figura 24**). Estos resultados nos sugieren que RUNX1 jugaría un rol dual en la regulación del elemento ERP, dicha regulación se caracteriza por disminuir la expresión de un gen reportero a bajas concentraciones de RUNX1 y aumentar dicha expresión al cotransfectar 500ng de RUNX1.





Fuente: Elaboración propia

Figura 24: La proteína RUNX1 modula la actividad transcripcional de la región ERP.

Células HeLa fueron cotransfectadas con 200 ng del vector ERP-Sentido pGL3-Basic junto a dosis crecientes del vector de expresión pRUNX1 que expresa la proteína RUNX1 (25 ng – 500ng), el ADN total de cada reacción fue completado a 751 ng de ADN total con el plásmido pBluescript SK(-). Todas las transfecciones fueron normalizadas con la cotransfección del plásmido de expresión de renilla y los datos son expresados como unidades relativa de luciferasa (URL). La actividad reportera fue cuantificada 18 horas post-transfección. Cada barra representa el promedio más el error estándar de la media (e.e.m) de cuatro experimentos independientes, cada uno realizado por triplicado. Los asteriscos representan la significancia estadística determinada por el *t-test* (***) $p < 0.001$, ns= no significativo).

DISCUSION

La OMS define a las leucemias como un proceso neoplásico caracterizado por la acumulación clonal de una célula tumoral, denominada blasto, en médula ósea y sangre periférica, que a su vez se caracterizan por la expresión de genes quiméricos que afectan los procesos de proliferación, maduración, diferenciación y apoptosis celular.

La identificación de anomalías genéticas asociadas al desarrollo de cáncer datan de la década del sesenta, donde Nowell, P y cols (Nowell 1962, Nowell 2007) utilizando técnicas de citogenética básica, identificó que el cariotipo presente en leucocitos aislados desde pacientes con Leucemia Mieloide Crónica (LMC) se caracterizaba por presentar un cromosoma pequeño, que solamente era identificado en los leucocitos de pacientes con la enfermedad y no así en los leucocitos aislados desde individuos normales. Este nuevo cromosoma fue denominado Cromosoma Filadelfia, en honor a la ciudad donde fue identificado. Posteriormente y gracias al desarrollo de técnicas citogenéticas más avanzadas Rowley y cols, determinaron que el cromosoma Filadelfia es el resultado de una translocación de tipo recíproca entre los cromosomas 9 y 22 y que es expresada específicamente en los leucocitos de pacientes con LMC (Rowley 1973). Rowley, ese mismo año también determinó que los leucocitos de algunos pacientes con leucemia mieloide aguda, y no así individuos normales, expresan la $t(8;21)$, lo cual le permitió sugerir que “Translocaciones cromosómicas específicas se asocian a enfermedades específicas” (Rowley

1973, Nucifora and Rowley 1995). Desde entonces se han reportado más de 10.000 tipos diferentes de eventos genéticos estructurales en 67.625 pacientes con distintos tipos de cáncer (Mitelman F 2017). De hecho, gracias a la asociación de antecedentes moleculares y clínicos, la OMS a partir del año 2002 ha determinado sellos genéticos recurrentes, que permiten el diagnóstico, monitoreo y evaluación pronóstica de pacientes con distintos tipos de leucemia (Vardiman, Harris et al. 2002, Arber, Orazi et al. 2016). A su vez estos estudios han permitido determinar que, de los aproximadamente 25.000 genes que conforman el genoma humano, solo un pequeño número es blanco frecuente de translocaciones cromosómicas y más aún, aparentemente existe cierta especificidad en cuanto a los compañeros de translocación para cada gen. (Mitelman, Johansson et al. 2007). De hecho, Mitelman y cols determinaron que existen genes altamente translocadores, los que pueden ser identificados en más de 50 translocaciones distintas, no obstante a pesar del alto número de genes compañeros de recombinación, las regiones de quiebre se localizan en regiones altamente específicas para cada compañero de translocación (Mitelman 2000, Mitelman, Johansson et al. 2007, Mitelman F 2017). Por ejemplo, el factor de transcripción RUNX1 ha sido identificado como uno de los genes maestros de la hematopoyesis y blanco frecuente de translocaciones cromosómicas (Levanon and Groner 2004). A la fecha, la proteína RUNX1 ha sido identificada en más de 60 tipos de translocaciones cromosómicas asociadas al desarrollo de neoplasias hematológicas. De hecho, estas observaciones han generado que la OMS, en su nueva clasificación de

neoplasias hematológicas basadas en sellos genéticos, distinga las neoplasias asociadas a alteraciones genéticas en el gen *RUNX1* como una entidad distinta al resto de las neoplasias debido a su factor pronóstico y expectativa de sobrevida de los pacientes afectados (Arber, Orazi et al. 2016).

Desde el punto de vista molecular, los eventos genéticos que afectan al gen *RUNX1* son principalmente translocaciones cromosómicas, donde las más frecuentes son las translocaciones cromosómicas (8;21), (12;21) y (3;21). En estas translocaciones, *RUNX1* recombina con los genes *ETO*, *EVI*, y *TEL*, respectivamente, en regiones específicas: el intrón 5 en la t(8;21), intrón 1 en t(12;21) e intrones 6 y 7 en t(3;21). Si bien el estudio de la expresión de los nuevos genes de fusión generados a partir de estos eventos de recombinación ha sido muy intensivo a nivel científico y clínico en los últimos cincuenta años, el mecanismo que explique los eventos de formación de la translocación o las causas de la especificidad de recombinación del gen *RUNX1* con su genes compañeros son desconocidos.

Hasta ahora, y a pesar de los esfuerzos realizados, sólo se han identificado algunas características de las zonas de quiebre cromosomal mapeadas en el locus de *RUNX1*: a) se localizan en regiones acotadas de los intrones, b) algunas de ellas presentan elementos estructurales de la cromatina como sitios de Hipersensibilidad a DNasa I, c) presentan patrones específicos de modificaciones de histonas, d) colocalizan o se encuentran próximas a regiones de corte para Topoisomerasa I. Algunas de estas características (b, c y d) han sido asociadas a la presencia de módulos *cis* reguladores. De hecho, la

búsqueda de sitios de Hipersensibilidad a DNasa I en conjunto con modificaciones de histonas específicas, son actualmente la estrategia de búsqueda mayormente utilizada en la identificación de nuevos elementos *cis* reguladores. Considerando estos antecedentes, en nuestro trabajo se formuló como hipótesis que *las regiones de quiebre cromosomal presentes en el locus del gen RUNX1 presentan módulos cis reguladores transcripcionalmente activos*. Para validar esta hipótesis y debido al tamaño del locus a analizar (260kb), realizamos la búsqueda de data libremente accesible y obtenida desde células hematopoyéticas mediante análisis de ChIPseq de las modificaciones de histonas H3K27ac, H3K4me3, H3K4me1 y la variante de histona H2A.Z. Se eligieron esas marcas debido a que su combinación, identifica de forma genérica al menos dos elementos *cis* reguladores: regiones promotoras y enhancers transcripcionalmente activos.

Al analizar primeramente la distribución de los peaks a lo largo del genoma, decidimos caracterizar su enriquecimiento en las regiones promotoras, intergénicas, exónicas e intrónicas, considerando a su vez su abundancia en pares de bases a nivel genómico, es decir, su porcentaje de representación a nivel de genoma completo. Nuestros resultados muestran que el enriquecimiento en H3K27ac fue mapeado preferentemente en regiones promotoras, lo que concuerda con la mayor parte de los trabajos, que describen la asociación genómica de dicha marca a regiones promotoras de genes que están siendo transcrita. No obstante, encontramos también un número importante de peaks enriquecidos en H3K27ac que fueron mapeados en regiones intragénicas, los

que probablemente indican potenciales enhancers o promotores alternativos (Cotney, Leng et al. 2012). Nuestro estudio evidenció que la mayor parte de las regiones intragénicas enriquecidas en H3K27ac corresponden a regiones intrónicas, lo que nos permite sugerir que en el genoma de las células GM12878 existen cerca de 20.000 elementos reguladores *cis* activos que a la fecha no han sido identificados.

De forma similar, al analizar el enriquecimiento genómico de H3K4me3, podemos observar una asociación preferente a regiones promotoras y exónicas, Estos resultados, también son concordantes con la literatura actual, debido a que se ha reportado que las regiones promotoras se encuentran enriquecidas en esta marca, debido posiblemente al rol de la H3K4me3 favoreciendo el reclutamiento de factores generales de la transcripción y por ende el inicio de la expresión génica (Vermeulen, Mulder et al. 2007, Zhou, Goren et al. 2011). Además, el enriquecimiento de las regiones exónicas identificadas puede ser explicado debido al rol de la modificación de histonas H3K4me3 en el proceso de *splicing*, ya que se ha reportado que esta marca se localiza también preferentemente en el sitio de unión entre el primer exón-intron, favoreciendo así el reconocimiento de la maquinaria de *splicing* (Bieberstein, Carrillo Oesterreich et al. 2012). También ha sido establecida la asociación de H3K4me3 a la ribonucleoproteína U2 y la RNA polimerasa II, favoreciendo el proceso de *splicing* cotranscripcional (Sims, Millhouse et al. 2007).

Por otra parte, y al igual que en la modificación anterior, también se determinó que la mayor proporción de las regiones promotoras y exónicas, se

encuentran enriquecidas en H3K4me1. En consiguiente, para esta marca epigenética, también se detectó una baja proporción de regiones intergénicas e intrónicas asociadas a H3K4me1, lo cual puede sugerir la presencia de eventuales *enhancers* distales, que presentan altos niveles de esta marca epigenética (Zhu, Sun et al. 2013, Ubhe, Rawat et al. 2017).

Por último, el análisis de la asociación de variante de histonas H2A.Z, demostró que H2A.Z se localiza preferentemente en regiones promotoras, resultado concordante con lo reportado previamente, donde su localización en regiones promotoras contribuiría a desestabilizar a la interacción entre el nucleosoma y el ADN, favoreciendo la unión de los factores de transcripción y el inicio de la expresión génica; aún así, se identificaron 19.653 regiones intrónicas enriquecidas en esta marca, lo cual puede sugerir que estas regiones son altamente accesible a la maquinaria transcripcional.

Posteriormente, al revisar el enriquecimiento de las marcas mencionadas a lo largo del locus de *RUNX1*, podemos evidenciar que algunas regiones presentes en intrones, exhiben enriquecimiento en algunas o todas las marcas epigenéticas evaluadas. Debido a que la presencia de estas marcas no siempre se asocia a actividad transcripcional, decidimos incorporar datos obtenidos mediante ensayos de hipersensibilidad a DNasa I acoplados a secuenciación masiva, como ayuda para evaluar cuáles de las regiones enriquecidas pueden presentar actividad transcripcional en el modelo evaluado (células linfoblásticas GM12878). Ahora bien, al utilizar la Hipersensibilidad a DNasa I como criterio primario de búsqueda, podemos destacar que el locus del gen *RUNX1* presenta

5 regiones con un alto grado de hipersensibilidad, a su vez estas regiones se caracterizan porque se localizan preferentemente en intrones donde han sido reportados quiebres cromosomales (el intrón 5 y 1) en translocaciones altamente recurrentes en clínica (t(8;21) y t(12;21) respectivamente).

Al analizar en detalle el intrón 5 podemos evidenciar que la región hipersensible a DNasa I y enriquecida en las marcas de histonas H3K4me3, H3K4me1, H3K27ac y H2A.Z colocaliza con una de las regiones de quiebre acotadas, descrita para la translocación (8;21). Lo cual, en este punto nos permitió sugerir que la región de quiebre cromosomal presente en el intrón 5 del gen RUNX1 presenta un elemento *cis* regulador, por esta razón denominamos a esta región como ERP (Elemento Regulador Putativo).

Considerando lo anterior decidimos clonar ERP en el vector pGL3 Basic y realizar transfecciones transientes en la línea celular HeLa. Nuestros resultados determinaron que ERP modula la expresión del gen reportero de luciferasa de manera orientación dependiente, lo cual nos permite concluir que la región ERP presenta actividad promotora *in vitro*. A su vez, es importante destacar que este promotor se localiza en el centro de una región de quiebre cromosomal. A la fecha, de todas las translocaciones y sitios de quiebre reportados, resultados similares sólo han sido descritos por Scharf y cols (Scharf, Zech et al. 2007) en una de las regiones de quiebre del gen *MLL*.

El gen *MLL* (Mixed Lineage Leukemia) codifica para una enzima de tipo metiltransferasa y se caracteriza por ser blanco frecuente de translocaciones cromosómicas, de hecho ha sido identificado en el 8% de las leucemias

linfoblásticas infantiles (Cimino, Rapanotti et al. 1998, Langer, Metzler et al. 2003). Este gen, al igual que el gen *RUNX1*, presenta múltiples compañeros de recombinación, uno de los sitios de quiebre mapeados, corresponde a una región de aproximadamente 8 Kb, en la que ha sido identificado un promotor interno a partir del cual se expresa un ARN con propiedades funcionales (Scharf, Zech et al. 2007). Nuestros análisis también determinaron que a partir del elemento ERP se expresa un ARN mensajero en células hematopoyéticas, lo que nos permite a su vez demostrar actividad transcripcional en la zona de quiebre.

Estos antecedentes resultan interesantes, ya que en los últimos años se ha establecido que regiones transcripcionalmente activas, como promotores, son más susceptibles a generar quiebres en la doble hebra de ADN (Aguilera and Gomez-Gonzalez 2008). De hecho, ya desde la década del ochenta, se ha descrito que de forma aparentemente normal la transcripción estimula la recombinación de tipo espontánea, en un fenómeno denominado “Recombinación Asociada a Transcripción” o RAT (Keil and Roeder 1984).

RAT fue primeramente observado en células eucariontes por Keil y cols, en *Saccharomyces cerevisiae* (Keil and Roeder 1984), donde demostró que el ADN ribosomal HOT1 es un sitio recurrente de recombinaciones cromosómicas en *S. cerevisiae*, pero dichas recombinaciones sólo pueden ocurrir en presencia de la ARN polimerasa y de transcripción activa. Recientemente también se ha establecido que existe una asociación entre el patrón de expresión génica, la estructura de cromatina presente en un promotor transcripcionalmente activo y la

formación de quiebres cromosomales. Esto gracias al desarrollo de nuevas técnicas de secuenciación masiva aplicadas al estudio de translocaciones cromosómicas, donde se ha determinado que utilizando como modelo de estudio el locus de c-myc y/o IgH, caracterizados ambos locus por presentar una cromatina altamente accesible y ser transcritos en células linfoides, los quiebres en la doble hebra de ADN ocurren casi exclusivamente en regiones transcripcionalmente activas (Janz 2006, Mathas, Kreher et al. 2009, Chiarle, Zhang et al. 2011).

De este modo, la transcripción activa parece ser un elemento fundamental en la formación de quiebres cromosomales. Estudios realizados por Mathas y cols, en células aisladas desde pacientes con Linfoma Anaplástico de Células Grandes evidenció que para que se produzca la translocación (2;5), es necesario primeramente un aumento en la expresión de los genes cercanos y que colocalizan con la región de quiebre para esta translocación (Mathas, Kreher et al. 2009).

Tomando en cuenta que en esta tesis se identificaron al menos 4 potenciales elementos reguladores transcripcionalmente activos localizados en regiones de quiebre cromosomal, donde uno de ellos presenta actividad promotora y promueve la expresión de un ARN mensajero, podemos sugerir que quizás las translocaciones cromosómicas que afectan al gen *RUNX1* se producen cuando regiones promotoras o enhancers, localizados en regiones intrónicas, se activan posiblemente de forma descontrolada, lo que contribuye a la formación de quiebres en la doble hebra de ADN en regiones específicas; es

decir, de cierta forma la formación de translocaciones cromosómicas se debe a un colapso de la burbuja transcripcional, lo cual podría también explicar la presencia de sitios de corte a Topoisomerasa II en las regiones de quiebre. La Topoisomerasa II es la enzima encargada de cortar el ADN y posteriormente volver a ligarlo para aliviar el estrés producido por la apertura de la doble hebra para posibilitar la elongación de la ARN polimerasa II. Si el proceso transcripcional es descontrolado, la producción de los quiebres mediados por Topoisomerasa II es mayor a la tasa de ligación de la misma enzima, favoreciendo de este modo la generación de quiebres en el ADN.

Resulta llamativo resaltar, que de acuerdo a nuestros resultados, la propia proteína RUNX1 controla la actividad transcripcional de la región ERP. A bajas concentraciones, RUNX1 presenta un efecto represor de la actividad transcripcional de ERP, mientras que a concentraciones elevadas de RUNX1 existe un aumento en la actividad transcripcional de ERP. Esto resulta interesante, ya que podría ayudar a entender, porque las translocaciones que afectan al gen *RUNX1* ocurren solo en células hematopoyéticas. Según nuestras observaciones, podemos proponer como respuesta que, debido a que en células hematopoyéticas existe una alta expresión de RUNX1, este activa al promotor ERP favoreciendo la formación de quiebres cromosomales. No obstante, en otro contexto celular no hematopoyético, la expresión de RUNX1 es baja o nula, por ende el promotor ERP no es activado y no se forman quiebres en esa región.

Para finalizar es importante destacar que este trabajo establece las bases para posteriores caracterizaciones funcionales de las regiones de quiebre

cromosomal. Como proyección, es necesario evaluar funcionalmente las regiones de quiebre en distintos contextos celulares, con el objetivo de determinar si existen otros elementos *cis* reguladores que no fueron identificados en la línea celular evaluada en este trabajo.



BIBLIOGRAFIA

- Abe, A., A. Katsumi, M. Kobayashi, A. Okamoto, M. Tokuda, T. Kanie, Y. Yamamoto, T. Naoe and N. Emi (2012). "A novel RUNX1-C11orf41 fusion gene in a case of acute myeloid leukemia with a t(11;21)(p14;q22)." *Cancer Genet* **205**(11): 608-611.
- Adamo, L. and G. Garcia-Cardena (2012). "The vascular origin of hematopoietic cells." *Dev Biol* **362**(1): 1-10.
- Aguilera, A. and B. Gomez-Gonzalez (2008). "Genome instability: a mechanistic view of its causes and consequences." *Nat Rev Genet* **9**(3): 204-217.
- Al-Sweedan, S. A., J. P. Neglia, M. E. Steiner, B. C. Bostrom, T. Casey and B. A. Hirsch (2007). "Characteristics of patients with TEL-AML1-positive acute lymphoblastic leukemia with single or multiple fusions." *Pediatr Blood Cancer* **48**(5): 510-514.
- Alarcon-Riquelme, M. E. (2003). "A RUNX trio with a taste for autoimmunity." *Nat Genet* **35**(4): 299-300.
- Alvarez-Venegas, R. and Z. Avramova (2005). "Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants." *Nucleic Acids Res* **33**(16): 5199-5207.
- Andreasson, P., M. Hoglund, A. N. Bekassy, S. Garwicz, J. Heldrup, F. Mitelman and B. Johansson (2000). "Cytogenetic and FISH studies of a single center

- consecutive series of 152 childhood acute lymphoblastic leukemias." Eur J Haematol **65**(1): 40-51.*
- *Andreoli, F. and A. Del Rio (2014). "Physicochemical modifications of histones and their impact on epigenomics." Drug Discov Today **19**(9): 1372-1379.*
 - *Arber, D. A., A. Orazi, R. Hasserjian, J. Thiele, M. J. Borowitz, M. M. Le Beau, C. D. Bloomfield, M. Cazzola and J. W. Vardiman (2016). "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia." Blood **127**(20): 2391-2405.*
 - *Arber, D. A., M. L. Slovak, L. Popplewell, V. Bedell, D. Ikle, J. D. Rowley, K. International Workshop on Leukemia and T. Prior (2002). "Therapy-related acute myeloid leukemia/myelodysplasia with balanced 21q22 translocations." Am J Clin Pathol **117**(2): 306-313.*
 - *Attarbaschi, A., G. Mann, M. Konig, M. N. Dworzak, M. M. Trebo, N. Muhlegger, H. Gadner, O. A. Haas and g. Austrian Berlin-Frankfurt-Munster cooperative study (2004). "Incidence and relevance of secondary chromosome abnormalities in childhood TEL/AML1+ acute lymphoblastic leukemia: an interphase FISH analysis." Leukemia **18**(10): 1611-1616.*
 - *Ausio, J., F. Dong and K. E. van Holde (1989). "Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome." J Mol Biol **206**(3): 451-463.*
 - *Bae, M. H., H. K. Kim, C. J. Park, E. J. Seo, S. H. Park, Y. U. Cho, S. Jang, H. S. Chi and K. H. Lee (2013). "A case of systemic mastocytosis associated with acute*

- myeloid leukemia terminating as aleukemic mast cell leukemia after allogeneic hematopoietic stem cell transplantation.* " Ann Lab Med **33**(2): 125-129.
- *Baron, M. H. and S. T. Fraser (2005). "The specification of early hematopoiesis in the mammal." Curr Opin Hematol **12**(3): 217-221.*
 - *Baron, M. H., J. Isern and S. T. Fraser (2012). "The embryonic origins of erythropoiesis in mammals." Blood **119**(21): 4828-4837.*
 - *Barrera, L. O. and B. Ren (2006). "The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding." Curr Opin Cell Biol **18**(3): 291-298.*
 - *Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.*
 - *Beck, D., J. A. Thoms, D. Perera, J. Schutte, A. Unnikrishnan, K. Knezevic, S. J. Kinston, N. K. Wilson, T. A. O'Brien, B. Gottgens, J. W. Wong and J. E. Pimanda (2013). "Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes." Blood **122**(14): e12-22.*
 - *Bieberstein, N. I., F. Carrillo Oesterreich, K. Straube and K. M. Neugebauer (2012). "First exon length controls active chromatin signatures and transcription." Cell Rep **2**(1): 62-68.*
 - *Blankenberg, D. and J. Hillman-Jackson (2014). "Analysis of next-generation sequencing data using Galaxy." Methods Mol Biol **1150**: 21-43.*

- Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey and G. E. Crawford (2008). "High-resolution mapping and characterization of open chromatin across the genome." *Cell* **132**(2): 311-322.
- Calero-Nieto, F. J., F. S. Ng, N. K. Wilson, R. Hannah, V. Moignard, A. I. Leal-Cervantes, I. Jimenez-Madrid, E. Diamanti, L. Wernisch and B. Gottgens (2014). "Key regulators control distinct transcriptional programmes in blood progenitor and mast cells." *EMBO J* **33**(11): 1212-1226.
- Chen, C. Y., Q. Morris and J. A. Mitchell (2012). "Enhancer identification in mouse embryonic stem cells using integrative modeling of chromatin and genomic features." *BMC Genomics* **13**: 152.
- Chiarle, R., Y. Zhang, R. L. Frock, S. M. Lewis, B. Molinie, Y. J. Ho, D. R. Myers, V. W. Choi, M. Compagno, D. J. Malkin, D. Neuberg, S. Monti, C. C. Giallourakis, M. Gostissa and F. W. Alt (2011). "Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells." *Cell* **147**(1): 107-119.
- Chinen, Y., T. Taki, K. Nishida, D. Shimizu, T. Okuda, N. Yoshida, C. Kobayashi, K. Koike, M. Tsuchida, Y. Hayashi and M. Taniwaki (2008). "Identification of the novel AML1 fusion partner gene, LAF4, a fusion partner of MLL, in childhood T-cell acute lymphoblastic leukemia with t(2;21)(q11;q22) by bubble PCR method for cDNA." *Oncogene* **27**(15): 2249-2256.
- Cimino, G., M. C. Rapanotti, T. Sprovieri and L. Elia (1998). "ALL1 gene alterations in acute leukemia: biological and clinical aspects." *Haematologica* **83**(4): 350-357.

- Cotney, J., J. Leng, S. Oh, L. E. DeMare, S. K. Reilly, M. B. Gerstein and J. P. Noonan (2012). "Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb." *Genome Res* **22**(6): 1069-1080.
- DeVilbiss, A. W., R. Sanalkumar, K. D. Johnson, S. Keles and E. H. Bresnick (2014). "Hematopoietic transcriptional mechanisms: from locus-specific to genome-wide vantage points." *Exp Hematol* **42**(8): 618-629.
- Dorigo, B., T. Schalch, K. Bystricky and T. J. Richmond (2003). "Chromatin fiber folding: requirement for the histone H4 N-terminal tail." *J Mol Biol* **327**(1): 85-96.
- du Preez, L. L. and H. G. Patterton (2013). "Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure." *Subcell Biochem* **61**: 37-55.
- Eissenberg, J. C. and A. Shilatifard (2010). "Histone H3 lysine 4 (H3K4) methylation in development and differentiation." *Dev Biol* **339**(2): 240-249.
- Foster, N., K. Paulsson, M. Sales, J. Cunningham, M. Groves, N. O'Connor, S. Begum, T. Stubbs, D. J. McMullan, M. Griffiths, N. Pratt and S. Tauro (2010). "Molecular characterisation of a recurrent, semi-cryptic RUNX1 translocation t(7;21) in myelodysplastic syndrome and acute myeloid leukaemia." *Br J Haematol* **148**(6): 938-943.
- Garcia-Ramirez, M., F. Dong and J. Ausio (1992). "Role of the histone "tails" in the folding of oligonucleosomes depleted of histone H1." *J Biol Chem* **267**(27): 19587-19595.

- Giguere, A. and J. Hebert (2010). "CLCA2, a novel RUNX1 partner gene in a therapy-related leukemia with t(1;21)(p22;q22)." *Cancer Genet Cytogenet* **202**(2): 94-100.
- Giguere, A. and J. Hebert (2011). "Microhomologies and topoisomerase II consensus sequences identified near the breakpoint junctions of the recurrent t(7;21)(p22;q22) translocation in acute myeloid leukemia." *Genes Chromosomes Cancer* **50**(4): 228-238.
- Giguere, A. and J. Hebert (2013). "Identification of a novel fusion gene involving RUNX1 and the antisense strand of SV2B in a BCR-ABL1-positive acute leukemia." *Genes Chromosomes Cancer* **52**(12): 1114-1122.
- Golub, T. R., G. F. Barker, S. K. Bohlander, S. W. Hiebert, D. C. Ward, P. Bray-Ward, E. Morgan, S. C. Raimondi, J. D. Rowley and D. G. Gilliland (1995). "Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia." *Proc Natl Acad Sci U S A* **92**(11): 4917-4921.
- Grunberg, S. and S. Hahn (2013). "Structural insights into transcription initiation by RNA polymerase II." *Trends Biochem Sci* **38**(12): 603-611.
- Gu, Z., M. Churchman, K. Roberts, Y. Li, Y. Liu, R. C. Harvey, K. McCastlain, S. C. Reshmi, D. Payne-Turner, I. Iacobucci, Y. Shao, I. M. Chen, M. Valentine, D. Pei, K. L. Mungall, A. J. Mungall, Y. Ma, R. Moore, M. Marra, E. Stonerock, J. M. Gastier-Foster, M. Devidas, Y. Dai, B. Wood, M. Borowitz, E. E. Larsen, K. Maloney, L. A. Mattano, Jr., A. Angiolillo, W. L. Salzer, M. J. Burke, F. Gianni, O. Spinelli, J. P. Radich, M. D. Minden, A. V. Moorman, B. Patel, A. K. Fielding, J. M. Rowe, S. M. Luger, R. Bhatia, I. Aldoss, S. J. Forman, J. Kohlschmidt, K. Mrozek, G. Marcucci,

- C. D. Bloomfield, W. Stock, S. Kornblau, H. M. Kantarjian, M. Konopleva, E. Paietta, C. L. Willman, M. L. Loh, S. P. Hunger and C. G. Mullighan (2016). "Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia." *Nat Commun* **7**: 13331.
- Guillemette, B., P. Drogaris, H. H. Lin, H. Armstrong, K. Hiragami-Hamada, A. Imhof, E. Bonneil, P. Thibault, A. Verreault and R. J. Festenstein (2011). "H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation." *PLoS Genet* **7**(3): e1001354.
- Gustafson, S. A., P. Lin, S. S. Chen, L. Chen, L. V. Abruzzo, R. Luthra, L. J. Medeiros and S. A. Wang (2009). "Therapy-related acute myeloid leukemia with t(8;21)(q22;q22) shares many features with de novo acute myeloid leukemia with t(8;21)(q22;q22) but does not have a favorable outcome." *Am J Clin Pathol* **131**(5): 647-655.
- Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenko, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." *Nature* **459**(7243): 108-112.
- Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford and B. Ren (2007). "Distinct and predictive chromatin signatures of

- transcriptional promoters and enhancers in the human genome." *Nat Genet* **39**(3): 311-318.
- Heinz, S. and C. K. Glass (2012). "Roles of lineage-determining transcription factors in establishing open chromatin: lessons from high-throughput studies." *Curr Top Microbiol Immunol* **356**: 1-15.
 - Heinz, S., C. E. Romanoski, C. Benner and C. K. Glass (2015). "The selection and function of cell type-specific enhancers." *Nat Rev Mol Cell Biol* **16**(3): 144-154.
 - Hosogane, M., R. Funayama, M. Shirota and K. Nakayama (2016). "Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body." *Cell Rep* **16**(3): 696-706.
 - Huang, B., C. Jiang and R. Zhang (2014). "Epigenetics: the language of the cell?" *Epigenomics* **6**(1): 73-88.
 - Janz, S. (2006). "Myc translocations in B cell and plasma cell neoplasms." *DNA Repair (Amst)* **5**(9-10): 1213-1224.
 - Jiao, L. and X. Liu (2015). "Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2." *Science* **350**(6258): aac4383.
 - Keil, R. L. and G. S. Roeder (1984). "Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*." *Cell* **39**(2 Pt 1): 377-386.
 - Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler and D. Haussler (2002). "The human genome browser at UCSC." *Genome Res* **12**(6): 996-1006.

- Kimura, H. (2013). "Histone modifications for human epigenome analysis." *J Hum Genet* **58**(7): 439-445.
- Kitabayashi, I., A. Yokoyama, K. Shimizu and M. Ohki (1998). "Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation." *EMBO J* **17**(11): 2994-3004.
- Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." *Science* **184**(4139): 868-871.
- Krauth, M. T., C. Eder, T. Alpermann, U. Bacher, N. Nadarajah, W. Kern, C. Haferlach, T. Haferlach and S. Schnittger (2014). "High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome." *Leukemia* **28**(7): 1449-1458.
- Langer, T., M. Metzler, D. Reinhardt, S. Viehmann, A. Borkhardt, M. Reichel, M. Stanulla, M. Schrappe, U. Creutzig, J. Ritter, T. Leis, U. Jacobs, J. Harbott, J. D. Beck, W. Rascher and R. Repp (2003). "Analysis of t(9;11) chromosomal breakpoint sequences in childhood acute leukemia: almost identical MLL breakpoints in therapy-related AML after treatment without etoposides." *Genes Chromosomes Cancer* **36**(4): 393-401.
- Levanon, D. and Y. Groner (2004). "Structure and regulated expression of mammalian RUNX genes." *Oncogene* **23**(24): 4211-4219.
- Li, B., S. G. Pattenden, D. Lee, J. Gutierrez, J. Chen, C. Seidel, J. Gerton and J. L. Workman (2005). "Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling." *Proc Natl Acad Sci U S A* **102**(51): 18385-18390.

- Li, L. and Y. Wang (2017). "Cross-talk between the H3K36me3 and H4K16ac histone epigenetic marks in DNA double-strand break repair." *J Biol Chem* **292**(28): 11951-11959.
- Lilljebjorn, H., R. Henningsson, A. Hyrenius-Wittsten, L. Olsson, C. Orsmark-Pietras, S. von Palffy, M. Askmyr, M. Rissler, M. Schrappe, G. Cario, A. Castor, C. J. Pronk, M. Behrendtz, F. Mitelman, B. Johansson, K. Paulsson, A. K. Andersson, M. Fontes and T. Fioretos (2016). "Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia." *Nat Commun* **7**: 11790.
- Luger, K., M. L. Dechassa and D. J. Tremethick (2012). "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?" *Nat Rev Mol Cell Biol* **13**(7): 436-447.
- Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 Å resolution." *Nature* **389**(6648): 251-260.
- Mahrez, W., M. S. Arellano, J. Moreno-Romero, M. Nakamura, H. Shu, P. Nanni, C. Kohler, W. Gruissem and L. Hennig (2016). "H3K36ac Is an Evolutionary Conserved Plant Histone Modification That Marks Active Genes." *Plant Physiol* **170**(3): 1566-1577.
- Mathas, S., S. Kreher, K. J. Meaburn, K. Johrens, B. Lamprecht, C. Assaf, W. Sterry, M. E. Kadin, M. Daibata, S. Joos, M. Hummel, H. Stein, M. Janz, I. Anagnostopoulos, E. Schrock, T. Misteli and B. Dorken (2009). "Gene deregulation and spatial genome reorganization near breakpoints prior to

- formation of translocations in anaplastic large cell lymphoma." *Proc Natl Acad Sci USA* **106**(14): 5831-5836.
- Mitelman, F. (2000). "Recurrent chromosome aberrations in cancer." *Mutat Res* **462**(2-3): 247-253.
 - Mitelman F, J. B. a. M. F. (2017). "Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer " <http://cgap.nci.nih.gov/Chromosomes/Mitelman>.
 - Mitelman, F., B. Johansson and F. Mertens (2007). "The impact of translocations and gene fusions on cancer causation." *Nat Rev Cancer* **7**(4): 233-245.
 - Nadal, N., J. L. Stephan, J. Cornillon, D. Guyotat, P. Flandrin and L. Campos (2008). "RUNX1 rearrangements in acute myeloblastic leukemia relapsing after hematopoietic stem cell transplantation." *Cancer Genet Cytogenet* **180**(2): 168-169.
 - Nekrasov, M., J. Amrichova, B. J. Parker, T. A. Soboleva, C. Jack, R. Williams, G. A. Huttley and D. J. Tremethick (2012). "Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics." *Nat Struct Mol Biol* **19**(11): 1076-1083.
 - Nekrasov, M., T. A. Soboleva, C. Jack and D. J. Tremethick (2013). "Histone variant selectivity at the transcription start site: H2A.Z or H2A.Lap1." *Nucleus* **4**(6): 431-438.
 - Nowell, P. C. (1962). "The minute chromosome (Ph1) in chronic granulocytic leukemia." *Blut* **8**: 65-66.

- Nowell, P. C. (2007). "Discovery of the Philadelphia chromosome: a personal perspective." *J Clin Invest* **117**(8): 2033-2035.
- Nucifora, G. and J. D. Rowley (1995). "AML1 and the 8;21 and 3;21 translocations in acute and chronic myeloid leukemia." *Blood* **86**(1): 1-14.
- Peeters, R., N. Buys, D. Vanmontfort, J. Van Isterdael, E. Decuypere and E. R. Kuhn (1992). "Preferential release of tri-iodothyronine following stimulation by thyrotrophin or thyrotrophin-releasing hormone in sheep of different ages." *J Endocrinol* **132**(1): 93-100.
- Pradeepa, M. M., G. R. Grimes, Y. Kumar, G. Olley, G. C. Taylor, R. Schneider and W. A. Bickmore (2016). "Histone H3 globular domain acetylation identifies a new class of enhancers." *Nat Genet* **48**(6): 681-686.
- Rabkin, C. S. and S. Janz (2008). "Overview of mechanisms and consequences of chromosomal translocation." *J Natl Cancer Inst Monogr*(39): 1.
- Rajagopal, N., J. Ernst, P. Ray, J. Wu, M. Zhang, M. Kellis and B. Ren (2014). "Distinct and predictive histone lysine acetylation patterns at promoters, enhancers, and gene bodies." *G3 (Bethesda)* **4**(11): 2051-2063.
- Rangasamy, D. (2013). "Distinctive patterns of epigenetic marks are associated with promoter regions of mouse LINE-1 and LTR retrotransposons." *Mob DNA* **4**(1): 27.
- Richardson, C. and M. Jasin (2000). "Frequent chromosomal translocations induced by DNA double-strand breaks." *Nature* **405**(6787): 697-700.
- Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G.

- Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn and W. J. Kent (2015). "The UCSC Genome Browser database: 2015 update." *Nucleic Acids Res* **43**(Database issue): D670-681.
- Rothbart, S. B. and B. D. Strahl (2014). "Interpreting the language of histone and DNA modifications." *Biochim Biophys Acta* **1839**(8): 627-643.
 - Rowley, J. D. (1973). "Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia." *Ann Genet* **16**(2): 109-112.
 - Roy, A. L. and D. S. Singer (2015). "Core promoters in transcription: old problem, new insights." *Trends Biochem Sci* **40**(3): 165-171.
 - Rye, M., P. Saetrom, T. Handstad and F. Drablos (2011). "Clustered ChIP-Seq-defined transcription factor binding sites and histone modifications map distinct classes of regulatory elements." *BMC Biol* **9**: 80.
 - Sakharkar, M. K., B. S. Perumal, K. R. Sakharkar and P. Kanguane (2005). "An analysis on gene architecture in human and mouse genomes." *In Silico Biol* **5**(4): 347-365.
 - Scharf, S., J. Zech, A. Bursen, D. Schraets, P. L. Oliver, S. Kliem, E. Pfitzner, E. Gillert, T. Dingermann and R. Marschalek (2007). "Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene." *Oncogene* **26**(10): 1361-1371.
 - Simon, M., J. A. North, J. C. Shimko, R. A. Forties, M. B. Ferdinand, M. Manohar, M. Zhang, R. Fishel, J. J. Ottesen and M. G. Poirier (2011). "Histone fold

- modifications control nucleosome unwrapping and disassembly." Proc Natl Acad Sci U S A **108**(31): 12711-12716.*
- *Sims, R. J., 3rd, S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage, P. Tempst, J. L. Manley and D. Reinberg (2007). "Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing." Mol Cell **28**(4): 665-676.*
 - *Smith, R. A. and C. A. Glomski (1982). ""Hemogenic endothelium" of the embryonic aorta: Does it exist?" Dev Comp Immunol **6**(2): 359-368.*
 - *Soboleva, T. A., M. Nekrasov, A. Pahwa, R. Williams, G. A. Huttley and D. J. Tremethick (2011). "A unique H2A histone variant occupies the transcriptional start site of active genes." Nat Struct Mol Biol **19**(1): 25-30.*
 - *Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.*
 - *Stuardo, M., M. Martinez, K. Hidalgo, M. Montecino, A. Javed, J. B. Lian, G. S. Stein, J. L. Stein and S. E. Gutierrez (2009). "Altered chromatin modifications in AML1/RUNX1 breakpoint regions involved in (8;21) translocation." J Cell Physiol **218**(2): 343-349.*
 - *Symmons, O. and F. Spitz (2013). "From remote enhancers to gene regulation: charting the genome's regulatory landscapes." Philos Trans R Soc Lond B Biol Sci **368**(1620): 20120358.*
 - *Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernet, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E.*

- Giste, A. K. Johnson, E. M. Johnson, T. Kutuyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford and J. A. Stamatoyannopoulos (2012). "The accessible chromatin landscape of the human genome." *Nature* **489**(7414): 75-82.
- Tse, C. and J. C. Hansen (1997). "Hybrid trypsinized nucleosomal arrays: identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction." *Biochemistry* **36**(38): 11381-11388.
- Ubhe, S., M. Rawat, S. Verma, K. Anamika and K. Karmodiya (2017). "Genome-wide identification of novel intergenic enhancer-like elements: implications in the regulation of transcription in *Plasmodium falciparum*." *BMC Genomics* **18**(1): 656.
- Vardiman, J. W., N. L. Harris and R. D. Brunning (2002). "The World Health Organization (WHO) classification of the myeloid neoplasms." *Blood* **100**(7): 2292-2302.
- Vardiman, J. W., J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, A. Porwit, N. L. Harris, M. M. Le Beau, E. Hellstrom-Lindberg, A. Tefferi and C. D. Bloomfield (2009). "The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes." *Blood* **114**(5): 937-951.

- Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann and H. T. Timmers (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." *Cell* **131**(1): 58-69.
- Vettese-Dadey, M., P. A. Grant, T. R. Hebbes, C. Crane- Robinson, C. D. Allis and J. L. Workman (1996). "Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro." *EMBO J* **15**(10): 2508-2518.
- Wang, S., C. Zang, T. Xiao, J. Fan, S. Mei, Q. Qin, Q. Wu, X. Li, K. Xu, H. H. He, M. Brown, C. A. Meyer and X. S. Liu (2016). "Modeling cis-regulation with a compendium of genome-wide histone H3K27ac profiles." *Genome Res* **26**(10): 1417-1429.
- Wiemels, J. L. and M. Greaves (1999). "Structure and possible mechanisms of TEL-AML1 gene fusions in childhood acute lymphoblastic leukemia." *Cancer Res* **59**(16): 4075-4082.
- Zhang, Y., P. Strissel, R. Strick, J. Chen, G. Nucifora, M. M. Le Beau, R. A. Larson and J. D. Rowley (2002). "Genomic DNA breakpoints in AML1/RUNX1 and ETO cluster with topoisomerase II DNA cleavage and DNase I hypersensitive sites in t(8;21) leukemia." *Proc Natl Acad Sci U S A* **99**(5): 3070-3075.
- Zhou, V. W., A. Goren and B. E. Bernstein (2011). "Charting histone modifications and the functional organization of mammalian genomes." *Nat Rev Genet* **12**(1): 7-18.

- Zhu, Y., L. Sun, Z. Chen, J. W. Whitaker, T. Wang and W. Wang (2013).
"Predicting enhancer transcription and activity from chromatin modifications."
Nucleic Acids Res **41**(22): 10032-10043.

