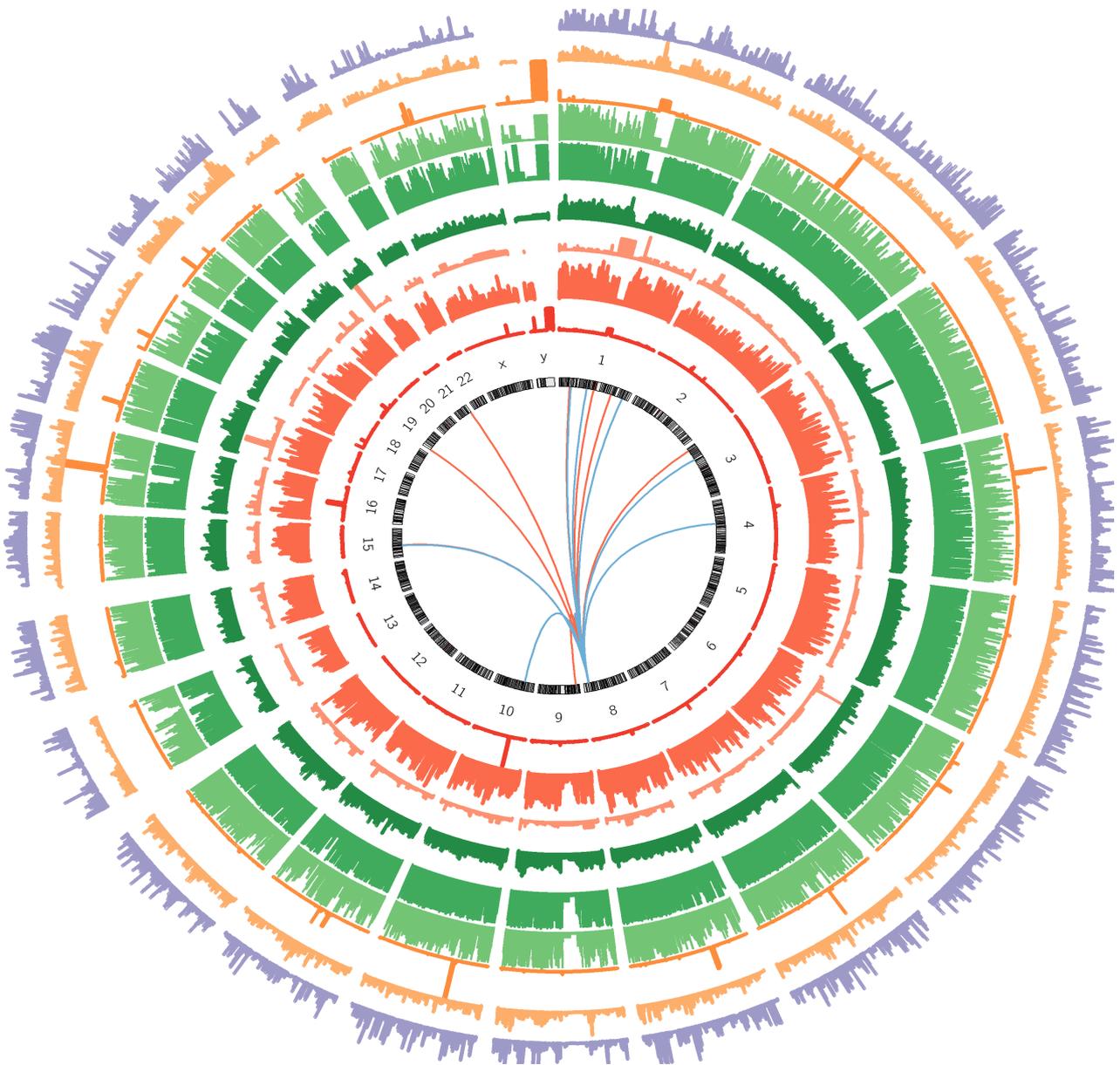


**Tesis de Maestría PEDECIBA**

**Área Biología – Subárea Biología Celular y Molecular**

**Análisis de las interacciones genómicas en una región del genoma  
(8q24) asociada al cáncer de próstata y colon, mediante captura de la  
conformación de la cromatina y secuenciación masiva**



**Programa de Desarrollo de las Ciencias Básicas (PEDECIBA)**

**Área Biología – Subárea Biología Celular y Molecular**

**Tesis de Maestría:**

**Análisis de las interacciones genómicas en una región del genoma (8q24)  
asociada al cáncer de próstata y colon, mediante captura de la  
conformación de la cromatina y secuenciación masiva**

Lic. Bioq. Valeria Romero

**Orientador:** José R. Sotelo-Silveira

**Co – orientador:** David Munroe

**Tribunal:**

Dr. Gustavo Folle (Presidente)

Dra. Beatriz Garat (Vocal)

Dr. Bernardo Bertoni (Vocal)

Instituto de Investigaciones Biológicas Clemente Estable

Diciembre 2014



## **Agradecimientos**

Al Coya, por aceptarme como estudiante e integrarme en sus proyectos, así como también brindarme su apoyo y confianza durante el transcurso de mi tesis.

A David, por recibirme en su laboratorio y ofrecerme todas las herramientas necesarias para desarrollarme .

A los miembros del Tribunal, por aceptar corregir mi tesis e interesarse en la misma.

A mis compañeros y amigos del DPAN y de Genómica, por acompañarme dentro y fuera del laboratorio.

A mi familia, muy especialmente a mis padres, por apoyarme y acompañarme desde siempre.

A mis amigos, que aunque nos dediquemos a cosas muy distintas o similares, y nos veamos poco, estuvieron presentes.

A mis compañeras de “La Española”, por su interés en mi trabajo y los consejos que me brindaron.

A Ariel, por acompañarme incondicionalmente y bancarme en todo momento.

A la ANII, por el apoyo económico, el cual me sustentó durante gran parte de mi trabajo.

# ÍNDICE

I. RESUMEN.....	1
II. INTRODUCCIÓN.....	3
8q24 y su asociación con cáncer.....	3
Sobreexpresión del gen c-Myc en cáncer.....	4
Elementos potenciadores en 8q24 y la regulación del gen c-Myc.....	5
Contactos físicos entre el promotor de c-Myc y el potenciador E en 8q24.....	7
Captura Circular de la Conformación de la Cromatina (4C) .....	9
Secuenciación masiva con Ion Torrent.....	10
III. HIPÓTESIS DE TRABAJO.....	12
IV. OBJETIVOS.....	13
Objetivo general.....	13
Objetivos específicos.....	13
V. MATERIALES Y MÉTODOS.....	14
Cultivos celulares.....	14
Circular Chromosome Conformation Capture o 4C.....	14
Entrecruzamiento.....	16
Digestión.....	16
Eficiencia de la digestión.....	17
Ligación.....	18
Reversión del entrecruzamiento.....	19
Purificación del ADN total.....	19
Purificación del ADN circular.....	19
Enriquecimiento por PCR anidado e invertido.....	20
Secuenciación de bibliotecas 4C.....	21
Selección de amplicones 4C.....	21
Optimización de la preparación de las librerías de secuenciación.....	21
Amplificación clonal por PCR en emulsión y secuenciación.....	21
Análisis de datos de secuenciación.....	22
Búsqueda de picos 4C.....	24
Análisis de datos 4C.....	25
Genes cercanos a picos 4C en común y a los más abundantes.....	25
Interpretación funcional.....	25
Intersección con datos de CHIP-Seq y DNAsa-Seq.....	26
Expresión en muestras clínicas.....	27
Literatura vinculada a cáncer.....	28

Visualización de datos e interacciones.....	28
Chromosome Conformation Capture o 3C.....	28
Entrecruzamiento.....	29
Digestión.....	29
Ligación.....	29
Reversión del entrecruzamiento.....	30
Purificación del ADN.....	30
Amplificación de moléculas conteniendo a las regiones que interactúan.....	30
Diseño de cebadores.....	30
Comparación con otros datos de interacciones en la línea LNCaP.....	35
VI. RESULTADOS.....	36
Generación de bibliotecas 4C.....	36
Enriquecimiento en moléculas conteniendo la carnada.....	37
Secuenciación de bibliotecas 4C.....	38
Optimización del tiempo de fragmentación.....	38
Librerías para secuenciar.....	39
Secuenciación en Ion Torrent.....	40
Análisis de datos de secuenciación.....	41
Búsqueda de picos 4C.....	46
Análisis de picos 4C.....	48
Distribución y posicionamiento de picos 4C en el genoma.....	48
Búsqueda de marcadores funcionales en coincidencia con las zonas de interacción obtenidas por 4C.....	52
Visualización de los picos 4C.....	59
Con IGV.....	59
Con Circos.....	60
Validación de contactos por 3C-PCR.....	63
Comparación con otros datos de interacciones en la línea LNCaP.....	67
VII. DISCUSIÓN.....	69
Producción de bibliotecas 4C conteniendo contactos genómicos para el potenciador E y c-Myc en LNCaP.....	69
La identificación de lecturas quiméricas con sitios de restricción para Mbo I indican las regiones ligadas a 8q24 en LNCaP .....	70
Picos 4C se corresponderían con los amplicones producidos por el PCR invertido y anidado.....	71
Los genes cercanos a los picos 4C compartidos entre el potenciador E y c-Myc tienen funciones relacionadas con desarrollo, muerte y sobrevivencia celular. ....	72
El mayor porcentaje de picos 4C de ambas carnadas se localizaron en el cromosoma 8.....	72

Regiones río abajo de genes que interactúan con el promotor de c-Myc podrían constituir potenciadores múltiples o “shadow enhancers” .....	73
Caracterización funcional de interactores por características de su cromatina.....	74
Estudio de marcadores de promotores y potenciadores.....	74
22 genes serían candidatos a interactuar con el potenciador E.....	74
94 potenciadores serían candidatos a interactuar con c-Myc.....	74
Genes y potenciadores candidatos (predichos por p300 y H3K4me1) muestran características de cromatina abierta, sugiriendo actividad transcripcional.....	75
Las cohesinas STAG1 y RAD21 mediarían, más que CTCF, los lazos de cromatina implicados en los contactos 4C identificados.....	76
Los sitios de TCF4 involucrados en los contactos 4C presentan características de cromatina abierta.....	78
Regiones funcionales seleccionadas como candidatos a interactuar con el potenciador E o promotor de c-Myc, presentan cromatina abierta y sitios de unión para cohesinas.....	78
Cada región estudiada puede interactuar con varias regiones genómicas.....	79
Comparación con otras metodologías del tipo 3C en LNCaP mostró algunas coincidencias. .	81
VIII. CONCLUSIONES Y PERSPECTIVAS.....	83
Conclusiones.....	83
Perspectivas.....	84
IX. BIBLIOGRAFÍA.....	85
Lista de Figuras.....	91
Lista de Tablas.....	92

## I. RESUMEN

La región 8q24 se caracteriza por presentar una baja densidad de genes; sin embargo, su amplificación en diversos tipos de cáncer y la observación de translocaciones cercanas a esta región, la convierten en un candidato interesante de estudio para dilucidar su rol biológico en los procesos tumorales. Recientemente se incrementó la atención sobre 8q24, ya que estudios de asociación del genoma asociaron varios polimorfismos con susceptibilidad para el desarrollo de tumores de colon y próstata.

La “cercanía” relativa del proto-oncogen c-Myc (a 330 kb), hizo especular que tal vez este gen estuviese bajo control de un potenciador situado en 8q24. Se demostró la existencia de dicho potenciador, y que su actividad es mediada por el complejo TCF4/ $\beta$ -catenina de la vía Wnt. Para que el potenciador en 8q24 regule a c-Myc, deberían estar próximos en el espacio nuclear. Estudios de captura de la conformación de la cromatina mostraron que existían lazos de cromatina, como consecuencia de su interacción a distancia, que los aproximaba. Este tipo de lazos es mediado por proteínas organizadoras del genoma, como CTCF y cohesinas. Adicionalmente en esta misma región, se descubrieron genes para ARNs largos no codificantes (lncRNAs) cerca o sobre el potenciador de 8q24, que afectan la expresión de c-Myc.

El objetivo de esta tesis fue investigar, utilizando una variante de los estudios de conformación de la cromatina, denominada Captura Circular de la Conformación de la Cromatina o 4C, la estructura de contactos y/o interacciones de largo alcance que establecen el potenciador en 8q24 y c-Myc con otras regiones del genoma. La técnica 4C implicó el entrecruzamiento de la cromatina, digestión del ADN unido por complejos proteicos, ligación del ADN para formar moléculas circulares y liberación de las mismas por reversión del entrecruzamiento. Estas moléculas contenían secuencias de regiones que estaban interactuando, y se enriqueció en aquellas que contenían al potenciador o a c-Myc, mediante PCR invertido y anidado con cebadores específicos. La identificación de los interactores se hizo a través de la secuenciación masiva de las moléculas circulares enriquecidas y el estudio bioinformático de las lecturas obtenidas usando como referencia el genoma de referencia humano anotado.

Una vez obtenida una lista de posibles puntos de contacto, se seleccionaron grupos de interactores (potenciadores de c-Myc o genes potenciados por la región 8q24) utilizando criterios funcionales. Específicamente se comparó los datos de 4C con datos de Inmunoprecipitación de la Cromatina para marcadores específicos de promotores y potenciadores (H3K4me1, H3K4me3, H3K27ac y p300), proteínas involucradas en la formación de los lazos de cromatina (CTCF, RAD21 y STAG1), el factor de transcripción (TCF4); y como marcadores de cromatina abierta o accesible, se utilizaron datos de inmunoprecipitación (ARN Polimerasa II) y de

hipersensibilidad a la DNAsa I. Adicionalmente se utilizaron de niveles de expresión en muestras clínicas, y se buscó literatura acerca de los vínculos de estas regiones con procesos tumorales.

Utilizando estos criterios se identificaron 7 genes que podrían ser regulados por el potenciador (ACBD3, BCL9, CCNY, HPSE, RTF1, SMARCC1, ZNF643), y 8 regiones que podrían actuar como potenciadores de c-Myc, (localizadas en 1p34.3, 1p11.2, 1q25.3, 3p25.3, 9p23, 15q15.3, 19p13.3 y 21q22.3). Algunos de ellos fueron confirmados mediante 3C-PCR y análisis de ADN quimérico en los datos 4C.

La literatura de los genes que interactúan con el potenciador E señala que algunos de ellos: a) están sobreexpresados en ciertos tipos de tumores; b) tienen relación con aspectos de la progresión y capacidad metastásica; c) por sus roles en la célula podrían influir en la transcripción de genes críticos para el control celular; d) están implicados en la regulación de la vía Wnt, de la cual c-Myc es un blanco.

Este trabajo permitió identificar los posibles interactores del potenciador E y del promotor de c-Myc en la línea LNCaP, a través del estudio de la conformación de la cromatina en combinación con secuenciación masiva. También se logró caracterizar funcionalmente dichas regiones, y se verificó su interacción a través de 3C-PCR.

A futuro, los interactores identificados podrían ser utilizados en análisis *in silico* y ensayos funcionales, para contribuir a la comprensión de sus roles biológicos en cáncer y a mayor escala, establecer las redes regulatorias que subyacen los procesos tumorales.

## II. INTRODUCCIÓN

### 8q24 y su asociación con cáncer

La región 8q24, localizada en el brazo largo del cromosoma 8 humano, comenzó a llamar la atención al observarse que estaba frecuentemente amplificada en diversos tipos de cáncer, tales como cáncer de próstata (Cher et al. 1996; Visakorpi et al. 1995; Sun et al. 2007; Ribeiro et al. 2006; Lapointe et al. 2007) y de colon (Douglas et al., 2004; Ried et al., 1996); o que estaba incluida en translocaciones causantes de otros tumores, como el linfoma de Burkitt (Battey et al. 1983; Toujani et al. 2009). 8q24 posee una baja densidad génica: secuencias retrovirales endógenas (HERVs) (Haiman et al. 2007), pequeños (Huppi et al. 2008) y largos ARNs no codificantes (lncRNAs) (Xiang et al. 2014; Ling et al. 2013; Yang et al. 2013; Chung et al. 2011)(Xiang et al. 2014; Ling et al. 2013) y pseudogenes y genes que se transcriben pero cuya función biológica se desconoce (DQ515897, CB104826, POU5F1P1) (Tomlinson et al. 2007) (Figura 1).

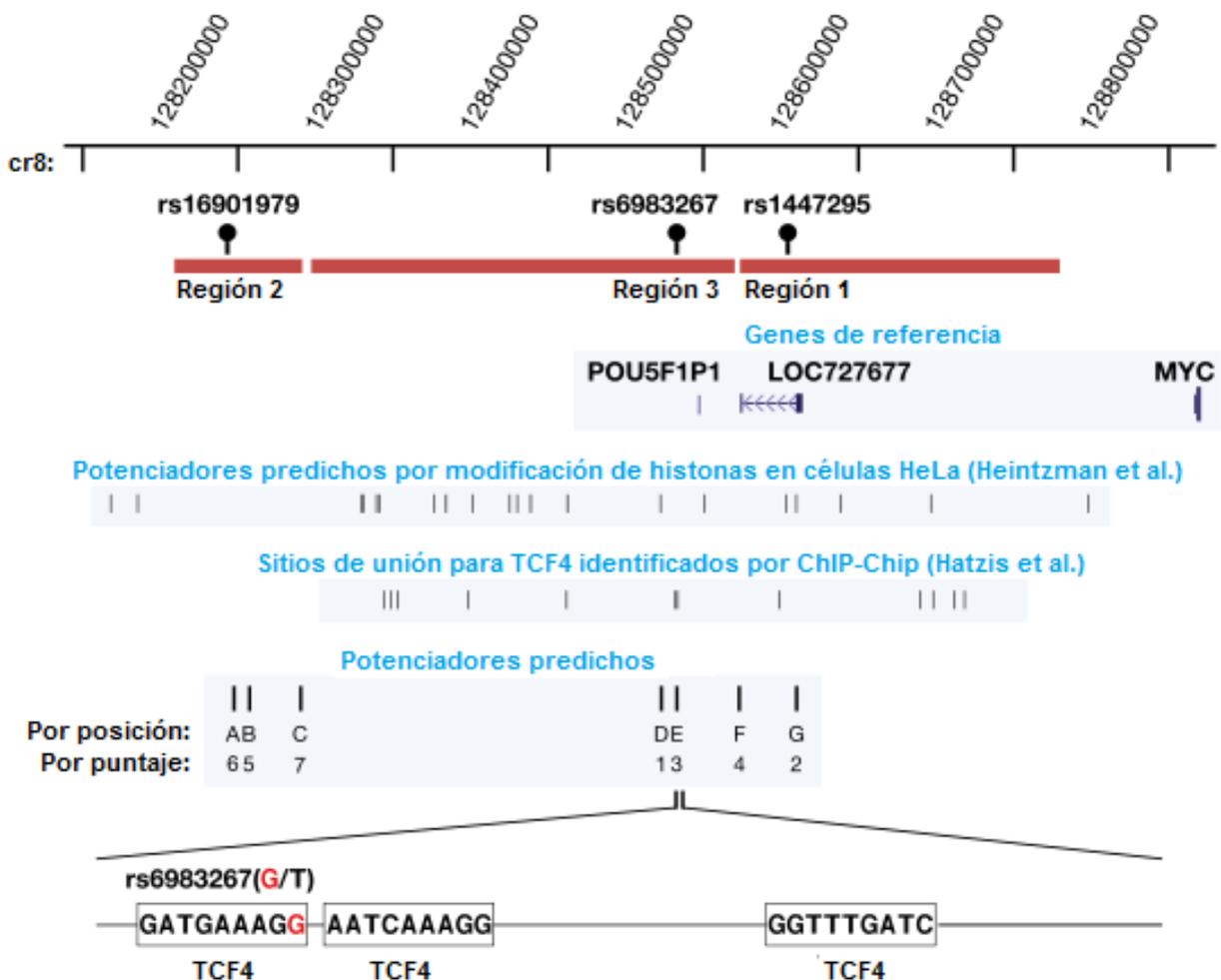


Figura 1 – Mapa físico de la región 8q24. En la parte superior se indican algunos SNPs asociados con cáncer, pseudogenes y genes, incluyendo al protooncogen c-Myc. En el medio se muestran elementos potenciadores propuestos y sitios de unión al factor de transcripción TCF4. En la parte inferior se indican elementos potenciadores identificados in silico (denominados como A – G), y se detallan los sitios de unión para TCF4 en el potenciador E (Modificado de Sotelo-Silveira et al. 2010).

Estudios de asociación del genoma o GWAS (Genome Wide Asociations Studies), mostraron que existían variantes o SNPs (Single Nucleotide Polymorphisms) en la secuencia de la región 8q24 asociadas a un mayor riesgo para cáncer de colon (Freedman et al. 2006; Tomlinson et al. 2007; Zanke et al. 2007) y próstata (Freedman et al. 2006; Haiman et al. 2007; Yeager et al. 2007; Gudmundsson et al. 2007; Witte 2007). Los GWAS permiten identificar variantes asociadas con susceptibilidad para una enfermedad sin ser necesario conocer *a priori* su rol (Easton and Eeles 2008; Couzin and Kaiser 2007).

La identificación de estos SNPs asociados permitió identificar tres bloques genómicos adyacentes en 8q24 que presentan un riesgo incrementado para cáncer de próstata (Witte 2007), y que fueron denominados regiones 1, 2 y 3 por el orden en que fueron descubiertos (Figura 1). El SNP rs6983267 localizado en la región 3, fue el que presentó una mayor significancia tanto para cáncer de colon como próstata. Este polimorfismo presenta dos alelos: T y G, siendo preferencialmente amplificado en cáncer de colon, el haplotipo que contiene al alelo G (Tuupanen et al. 2009).

Otros SNPs que también fueron asociados son rs1447295 y rs16901979 con cáncer de próstata (Witte 2007); rs10808556 y rs7013278 con colon (Haiman et al. 2007); rs9642880 con cáncer de vejiga (Gudmundsson et al. 2007) y rs13281615 con cáncer de mama (Easton et al. 2007).

### **Sobreexpresión del gen c-Myc en cáncer**

A una distancia promedio de 330 Kb de rs6983267 en 8q24, se encuentra el protooncogen c-Myc, implicado en procesos celulares tales como control del crecimiento celular, diferenciación celular, y apoptosis (Wierstra and Alves 2008). Mutaciones en la secuencia codificante de este gen no estarían exclusivamente vinculadas con la carcinogénesis (Levens 2010), pero sí su sobreexpresión (Buttayan et al. 1987; Gil et al. 2005; Williams et al. 2005).

Se han propuesto distintos mecanismos responsables de la sobreexpresión del gen c-Myc. Uno de ellos podría ser el aumento en el número de copias del gen como consecuencia de la amplificación observada, en ciertas patologías, de 8q24. Sin embargo, 8q24 sería predominantemente amplificada en tumores de próstata en etapa avanzada (Van Den Berg et al. 1995), mientras que la sobreexpresión del gen c-Myc ocurriría en tumores en etapa temprana (Gurel et al. 2008). Además, no se encontró correlación entre la amplificación de 8q24 confirmada por Hibridización *In Situ* por Fluorescencia o FISH (Fluorescence In Situ Hybridization) y los niveles de proteínas c-MYC en cáncer de próstata (Gurel et al. 2008).

Otro mecanismo de regulación propuesto de regulación es a través de la vía de señalización Wnt, implicada en procesos de desarrollo y cáncer, entre cuyos genes blanco se encuentra c-Myc. En esta vía participa la  $\beta$ -catenina, un cofactor de transcripción cuya activación en próstata de ratón resulta en niveles muy elevados de c-Myc y causa PIN (Prostatic Intraepithelial Neoplasia) (Yu et al. 2009). Otra opción es Foxp3, un factor de transcripción localizado en el cromosoma X que reprime al gen c-Myc. La “downregulation” de Foxp3 resulta en niveles mayores de ARNm y proteínas c-MYC en cáncer de próstata (Wang et al. 2009; Shen and Abate-shen 2010).

Como c-Myc es el protooncogen situado más cercano a la región 8q24 asociada a cáncer, y que previamente se sugirió que su expresión podría ser controlada por elementos distales (Lavenu et al. 1994; Hallikas et al. 2006), parte de las investigaciones previas a esta tesis se centró en la existencia de elementos reguladores de tipo “enhancer” o potenciadores en 8q24, que actúen sobre el gen c-Myc (Haiman et al. 2007).

### **Elementos potenciadores en 8q24 y la regulación del gen c-Myc**

Estudios recientes utilizaron la herramienta computacional EEL (Enhancer Element Locator) (Smith et al. 1981) para predecir elementos reguladores en la región 8q24 (Tuupanen et al. 2009; Sotelo et al. 2010). EEL predice elementos potenciadores evolutivamente conservados, basado en alineamientos de sitios de unión para factores de transcripción. Ambos trabajos coincidieron en la existencia de un elemento potenciador (denominado potenciador E) en la región donde se localiza rs6983267 (Tuupanen et al. 2009; Sotelo et al. 2010).

La función biológica de este elemento potenciador se comprobó a través de experimentos de transactivación *in vitro* en líneas celulares derivadas de cáncer de próstata y colon (heterocigotas para rs6983267), con un vector de expresión que contenía a dicho elemento junto al promotor mínimo del gen c-Myc (Sotelo et al. 2010) o al promotor del gen timidina quinasa (Tuupanen et al. 2009; Pomerantz et al. 2009), y al gen de la luciferasa como gen reportero de la transactivación transcripcional.

Ensayos de inmunoprecipitación de la cromatina o ChIP (Chromatin ImmunoPrecipitation) demostraron que el elemento potenciador E se une al factor de transcripción TCF4 (Tuupanen et al. 2009; Pomerantz et al. 2009; Sotelo et al. 2010), que también participa en la vía Wnt. La variante rs6983267 mapea en la última posición de la secuencia consenso de uno de los sitios de unión para TCF4 en la región de 8q24 donde se localiza el potenciador E (Hallikas et al. 2006; Tuupanen et al. 2009) (Figura 1). Aunque se observó que el alelo G de dicha variante

incrementa considerablemente la afinidad de TCF4 por este sitio (respecto al alelo T), no se encontró una correlación entre el genotipo rs6983267 y la expresión del gen c-Myc (Tuupanen et al. 2009; Pomerantz et al. 2009).

En los ensayos de ChIP también se observó coprecipitación de  $\beta$ -catenina (Tuupanen et al. 2009), cofactor de TCF4, lo que dio lugar a que experimentos de cotransfección con vectores de expresión para  $\beta$ -catenina y para TCF4, demostraran que la actividad del potenciador E depende de la coexpresión de estos factores (Sotelo et al. 2010). La actividad del potenciador E en 8q24 también se demostró *in vivo*, mediante la utilización de cromosomas artificiales de bacterias o BAC (Bacterial Artificial Chromosome) que contenían la secuencia del potenciador, el promotor mínimo de la  $\beta$ -globina y el gen de la  $\beta$ -galactosidasa como reportero (Wasserman et al. 2010).

Trabajos recientes reportaron la expresión de lncRNAs en la región 8q24 en cáncer de próstata (PCa) y de colon-recto (CRC) (revisados en Deng et al. 2014). En PCa se identificó al lncRNA PRNCR1 (Chung et al. 2011); mientras que en CRC se identificaron tres: CCAT1 (Nissan et al. 2012), CCAT1-L (Xiang et al. 2014), y CCAT2 (Ling et al. 2013).

En CRC se observó una asociación significativa entre el lncRNA CCAT2 y c-Myc. CCAT2 se localiza a 335 kb río arriba de c-Myc, sobre el rs6983267 en el potenciador E de 8q24 (Ling et al. 2013). CCAT2 potencia a TCF4 lo que provoca una activación aberrante de la vía Wnt y sobreexpresión de c-Myc. Esto se confirmó al realizar un knockdown de este lncRNA que redujo los niveles de expresión del ARN y de la proteína de c-Myc (Ling et al. 2013). A su vez, ensayos de interferencia para el ARNm de TCF4 mostraron una reducción en la expresión de CCAT2, sugiriendo la existencia de un mecanismo de retroregulación (Ling et al. 2013).

También se observó una asociación entre los niveles de expresión de CCAT2 con genotipo GG para rs6983267 y los niveles de c-Myc, sugiriendo que los alelos de rs6983267 podrían afectar diferencialmente la expresión de c-Myc (Ling et al. 2013), contrario a lo mencionado anteriormente (Tuupanen et al. 2009; Pomerantz et al. 2010) Respecto al mecanismo de regulación de c-Myc, CCAT2 parecería ser una vía alternativa a la del potenciador E. Sin embargo, cabe la posibilidad de que CCAT2 participe en la formación del lazo entre rs6983267 y el promotor de c-Myc, y que colaboren con el potenciador E para regular la expresión de este gen (Ling et al. 2013).

Además del elemento potenciador E en 8q24, se conoce un potenciador 1.6 kb río abajo de c-Myc (He et al. 1998; Yochum et al. 2007, 2008) y sitios de unión para decenas de factores de transcripción que modulan la expresión del gen c-Myc, situados río arriba de su sitio de inicio de

la transcripción o TSS (Transcription Start Site) (Wierstra and Alves 2008). Por otro lado, la presencia de muchos sitios de unión para factores de transcripción en el potenciador E llevó a proponer a esta región como un superpotenciador (Lovén et al. 2013). El hallazgo de estos nuevos potenciadores sugiere que quizás existan más por descubrir.

### Contactos físicos entre el promotor de c-Myc y el potenciador E en 8q24

Para que un gen interactúe con elementos reguladores tipo potenciador, deben estar próximos entre sí dentro de la organización tridimensional de la cromatina. Esto implica que, en un núcleo interfásico, los cromosomas no se posicionen al azar, sino relativos unos a otros para que la interacción sea reproducible en cada nueva instancia de regulación (Fraser and Bickmore 2007; Misteli 2007).

Las interacciones de tipo “long-range” o de largo alcance implican el contacto físico entre regiones distales en el mismo (*cis*) o distinto cromosoma (*trans*), y tiene como consecuencia la formación de “loops” o lazos de la cromatina que acercan dichas regiones (Figura 2). Ejemplos de interacciones de este tipo son la regulación de varios genes distales (incluso de distintos cromosomas) por potenciadores (Spilianakis et al. 2005; West and Fraser 2005), y la colocalización de genes transcripcionalmente activos y ARN polimerasas asociadas a sus promotores, en sitios del núcleo conocidos como “factories” o fábricas de transcripción (Xu and Cook 2008; Osborne et al. 2004; Müller et al. 2007; Pombo et al. 1999).

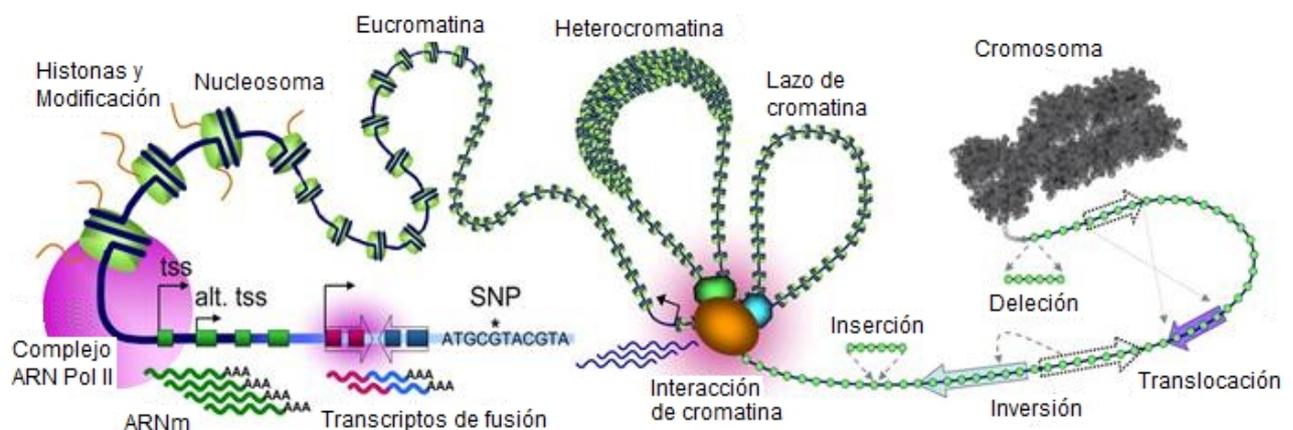


Figura 2 – Interacciones de largo alcance. La expresión de un gen puede controlarse a través del nivel de compactación de la cromatina y de la interacción con elementos reguladores. Las interacciones de largo alcance implican el contacto físico entre regiones distales en el mismo (*cis*) o distinto cromosoma (*trans*), y tiene como consecuencia la formación de “loops” o lazos de cromatina que acercan dichas regiones (Modificado de Fullwood et al. 2009).

Además de ARN polimerasas y otros componentes de la maquinaria transcripcional, los lazos de cromatina pueden ser mediados por la unión e interacción entre el factor CTCF y el complejo de cohesinas, considerados organizadores globales del genoma (Phillips and Corces 2009; Wendt et al. 2008).

CTCF fue inicialmente estudiado por su capacidad de interferir con la comunicación entre un potenciador y su promotor diana y establecer dominios funcionales de expresión génica (Bell et al. 1999; Kim et al. 2007; Handoko et al. 2011). Sin embargo, recientes trabajos en células madre embrionarias de ratón o ESCs (Embryonic Stem Cells), revelaron un rol primario de CTCF en mediar los lazos de cromatina y promover la comunicación entre promotores y potenciadores distales, más que bloquearla (Handoko et al. 2011; Dekker et al. 2014; Sanyal et al. 2012).

Las cohesinas son un complejo heterodímero constituido por cuatro subunidades (SMC1A, SMC3, RAD21 y STAG (1, 2 o 3)); y están implicadas en la cohesión de las cromátidas hermanas durante la mitosis y meiosis (Wendt et al. 2008). Las cohesinas colocalizan con CTCF, sugiriendo una función facilitadora en la formación de los lazos de cromatina (Dorsett 2011; Parelho et al. 2008), aunque también se han observado sitios de unión independientes de CTCF (Faure et al. 2012; Rubio et al. 2008).

Para estudiar las interacciones de largo alcance y conocer un poco mejor la conformación de la cromatina en dichas interacciones, se puede emplear la metodología 3C (Chromosome Conformation Capture) (Dekker et al. 2002). 3C pone en evidencia contactos físicos entre dos regiones genómicas que pueden localizarse en el mismo cromosoma o en distintos cromosomas, midiendo la frecuencia con la que suceden dichos contactos mediante PCR cuantitativo (Polymerase Chain Reaction) (Figura 3). Por este motivo, la metodología 3C fue utilizada por distintos grupos de investigación para evidenciar la interacción entre el potenciador E de 8q24 y el promotor y primera mitad del gen c-Myc en cáncer de colon y próstata (Pomerantz et al. 2009; Sotelo et al. 2010).

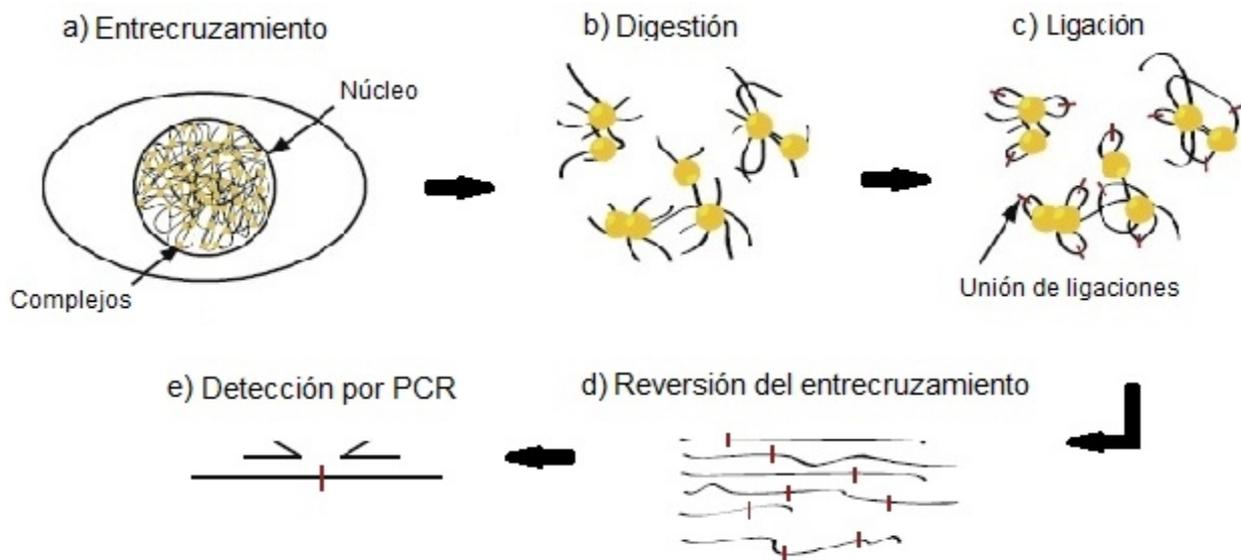


Figura 3 – Esquema de la metodología 3C. a) Entrecruzamiento de la cromatina para formar complejos ADN – proteínas que estén cercanos en el núcleo; b) digestión enzimática; c) ligación de los fragmentos de restricción unidos por el entrecruzamiento; d) reversión del mismo y e) detección de las uniones entre los fragmentos por PCR con cebadores específicos de las regiones de interés que flanqueen un sitio de restricción (Modificado de Naumova et al. 2012).

### Captura Circular de la Conformación de la Cromatina (4C)

Todos los estudios realizados sobre 8q24 indican que el elemento regulador que contiene al rs6983267, se trata de un potenciador que efectivamente actúa sobre un gen situado lejano (*c-Myc*), independientemente de su orientación tanto *in vivo* como *in vitro* (Tuupanen et al. 2009; Pomerantz et al. 2009; Sotelo et al. 2010), mediado por factores (Sotelo et al. 2010) y establece contactos físicos con regiones genómicas distales (Pomerantz et al. 2009; Sotelo et al. 2010).

El descubrimiento de nuevos elementos potenciadores en la región 8q24 motivó a pensar que existen más contactos por caracterizar. Como la metodología 3C permite solamente verificar un par de interactores a la vez, esta tesis propone utilizar la metodología Captura Circular de la Conformación de la Cromatina o 4C (Circular Chromosome Conformation Capture) (Zhao et al. 2006), la cual deriva del concepto de captura de la conformación de la cromatina pensado para 3C, para estudiar las interacciones a largo alcance de 8q24 a nivel genómico.

La metodología 4C permite capturar los contactos entre una región de interés (“bait” o carnada) y otras regiones de ADN próximos en el espacio nuclear, midiendo la frecuencia de tales contactos. Esta captura se logra fijando regiones de ADN que interactúan, con las proteínas que median dicha interacción (“crosslinking” o entrecruzamiento), y puede evidenciarse por secuenciamiento masivo con cebadores específicos para la carnada (Figura 4).

El procedimiento implica: a) entrecruzamiento de la cromatina para formar complejos ADN – proteínas que estén cercanos en el núcleo; b) digestión enzimática del ADN; c) ligación de los fragmentos de restricción unidos por el entrecruzamiento; d) reversión del mismo para generar una biblioteca 4C: moléculas circulares que contienen regiones genómicas que posiblemente estuviesen interactuando; y e) enriquecimiento de la biblioteca 4C en moléculas que contengan a la carnada por PCR invertido y anidado con cebadores específicos para la misma (Figura 4).

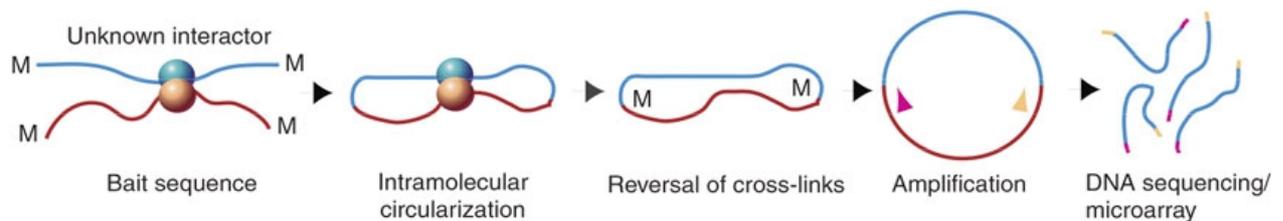


Figura 4 – Esquema de la metodología 4C. En color rojo oscuro se indica la secuencia de la carnada, y en color azul las secuencias desconocidas que interactúan con ella. Las cabezas de flecha indican la posición y dirección de los cebadores del PCR invertido y anidado (extraído de Göndör et al. 2008).

Los productos del PCR corresponden a las regiones genómicas con las que interactúa la carnada. La identificación de dichas regiones puede hacerse mediante secuenciación masiva y alineamiento contra un genoma de referencia, o mediante microarreglos con sondas conocidas del genoma.

### Secuenciación masiva con Ion Torrent

Para secuenciar en el equipamiento Ion Torrent (Life Technologies) se prepara una librería de secuenciación: se lleva la muestra a un tamaño esperado por fragmentación enzimática o sonicación y se ligan adaptadores en los extremos de las moléculas obtenidas, que pueden contener una pequeña secuencia conocida como código de barra para discriminar entre varias librerías, si se desea secuenciar en conjunto. Cuando se fragmentan grandes moléculas, el rango del tamaño final es muy variable, por lo que se lo puede acotar por distintos métodos.

Los adaptadores permiten la amplificación clonal de las moléculas en partículas esféricas o ISPs (Ion Spheres Particles) mediante un PCR en emulsión (emPCR). Las ISPs contienen en su superficie una secuencia complementaria a uno de los adaptadores de la librería de secuenciación, por lo que después del emPCR, cada esfera termina “cubierta” de copias de moléculas. Es importante cuantificar las librerías previo al emPCR, para mezclarlas en una relación adecuada con las ISPs y que éstas resulten monoclonales. Si la relación está desbalanceada, se pueden obtener partículas vacías (sin señal) o policlonales (con copias de dos o más moléculas, múltiples señales).

Luego se controla la calidad del emPCR, para estimar si fue suficiente y mayoritariamente monoclonal para secuenciar. El uso de sondas específicas para los adaptadores permite calcular la relación entre las moléculas amplificadas y las ISPs en una planilla excel proporcionada por el fabricante. Después se procede con el enriquecimiento en partículas que contienen molde, removiendo las vacías, aunque no se discrimina entre mono o policlonales.

Finalmente se carga la librería en un chip semiconductor, junto con la polimerasa y los cebadores para la secuenciación, donde se da el secuenciado por síntesis de ADN. El secuenciador introduce en el chip un nucleótido por vez. Cuando el nucleótido es incorporado a la hebra nascente de ADN, se libera un hidrogenión, cambia el pH de la solución y se produce una diferencia de potencial o voltaje, lo cual es detectado por el equipo, asignando la base correspondiente.

Si el nucleótido no se incorpora, no habrá cambio de voltaje y no se asignará ninguna base para ese punto en ese microrreactor. Si hay dos bases idénticas consecutivas el voltaje será el doble y se asignarán dos bases.

### **III. HIPÓTESIS DE TRABAJO**

Dado que la región 8q24: a) está amplificada en distintos tipos de cáncer e involucrada en translocaciones causantes de otros tipos; b) presenta al menos un elemento potenciador que afecta al gen c-Myc (localizado aproximadamente a 330 Kb); c) que c-Myc también estaría regulado por otros elementos potenciadores, la hipótesis de trabajo de esta Tesis es que la región 8q24 posee características estructurales particulares que le permiten influenciar la actividad génica relacionada a procesos de crecimiento celular anormal y que esta región y el gen c-Myc pueden establecer interacciones con regiones del genoma que se encuentren lejanas posicionalmente, posiblemente involucradas en patologías como el cáncer.

## **IV. OBJETIVOS**

### **Objetivo general**

Identificar las interacciones que establece la región 8q24 en la línea celular LNCaP, para contribuir a la comprensión de su rol biológico en cáncer de próstata.

### **Objetivos específicos**

- I. Analizar interacciones del potenciador E (que contiene al rs6983267 asociado a cáncer) en 8q24, mediante captura de la conformación de la cromatina (4C), para conocer cuáles regiones del genoma se encontrarían posiblemente bajo control de este elemento regulador.
- II. Analizar las interacciones del promotor de c-Myc, mediante captura de la conformación de la cromatina (4C), para conocer cuáles regiones del genoma regularían a este gen.
- III. Confirmar las interacciones identificadas por 4C mediante con 3C-PCR.

## **V. MATERIALES Y MÉTODOS**

### **Cultivos celulares**

Se cultivó la línea celular LNCaP en medio RPMI 1640 suplementado con 10% de suero bovino fetal (FBS), a 37°C y una atmósfera 5% de CO<sub>2</sub>. LNCaP es una línea derivada de una metástasis de cáncer de próstata y heterocigota para el SNP rs6983267; es hipotetraploide, crece en monocapa y es dependiente de andrógenos (Horoszewicz et al. 1983).

### **Circular Chromosome Conformation Capture o 4C**

Las interacciones genómicas del potenciador E y del promotor de c-Myc se identificaron mediante la metodología 4C (Zhao et al. 2006), de acuerdo al protocolo utilizado por Göndör y colaboradores (Göndör et al. 2008).

El procedimiento implicó: a) entrecruzamiento de la cromatina para formar complejos ADN – proteínas que estuviesen cercanos en el núcleo; b) digestión enzimática del ADN y control de la eficiencia de la digestión; c) ligación de los fragmentos de restricción unidos por el entrecruzamiento; d) reversión del mismo para generar una biblioteca 4C, compuesta por moléculas circulares que contengan regiones genómicas que posiblemente estuviesen interactuando; e) purificación del ADN circular; y f) enriquecimiento de la biblioteca 4C en moléculas que contengan a la carnada, por PCR invertido y anidado con cebadores específicos para la misma (Figura 4).

Por cada experimento 4C se utilizaron tres viales: 1 – muestra 4C: siguió todo el procedimiento y fue la utilizada para identificar los interactores; 2 – control ligado: siguió todo el procedimiento con excepción de que la reversión del entrecruzamiento fue anterior a la ligación, por lo tanto los fragmentos de restricción se liberaron de los complejos y se ligaron sobre sí mismos; 3 – control no ligado: siguió todo el procedimiento pero al final no se ligó (Figura 5).

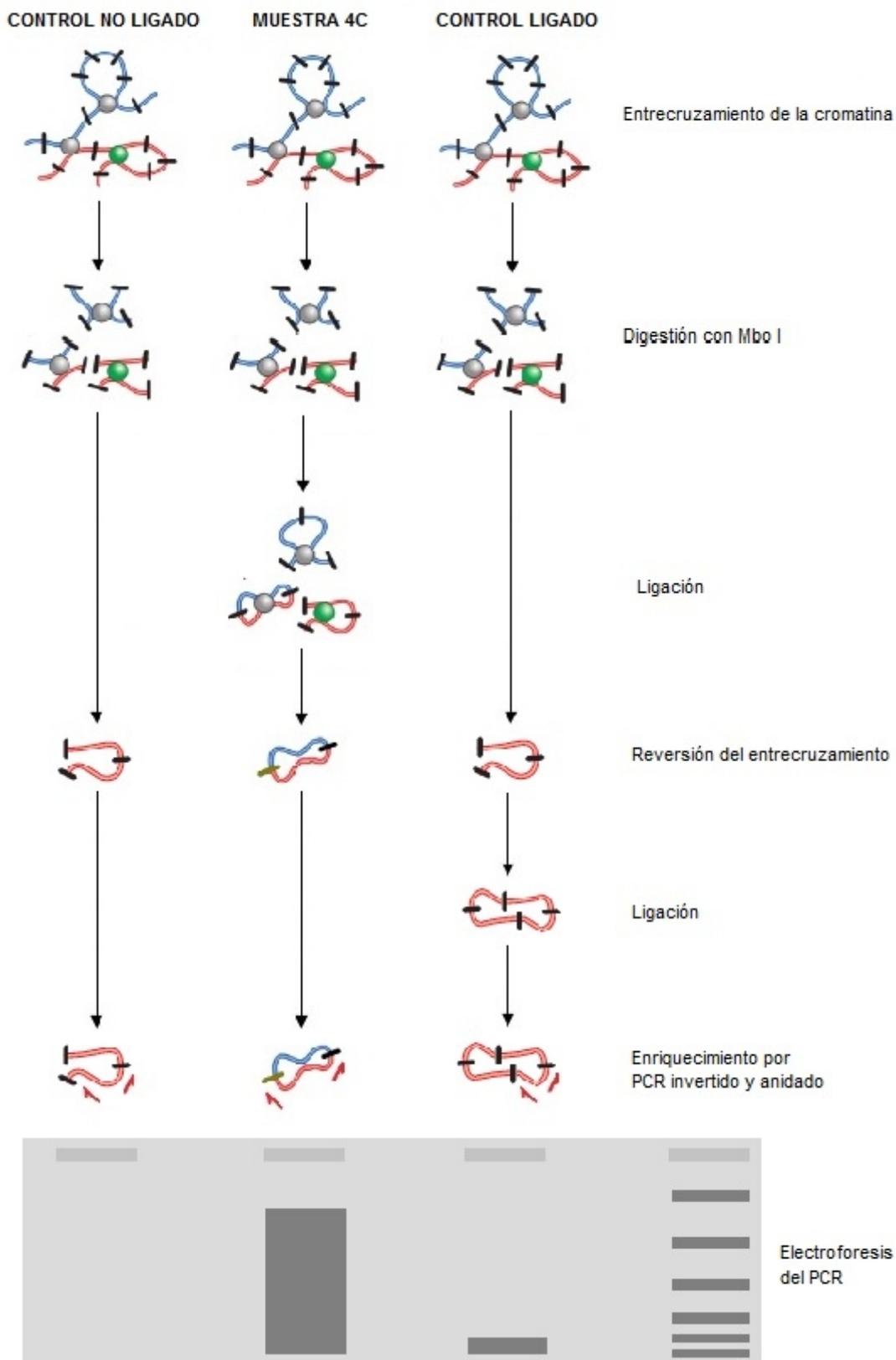


Figura 5 – Metodología 4C en detalle. En color rojo se indica la secuencia de interés o carnada, y en color azul la de las secuencias interactoras (modificado de van Steensel & Dekker, 2010).

## Entrecruzamiento

Cuando las células LNCaP alcanzaron un 80% de confluencia, se procedió con el entrecruzamiento de la cromatina. Se removió el medio de cultivo, se lavaron las células con PBS a temperatura ambiente y se incubaron con paraformaldehído al 1% en PBS (preparado fresco), durante 10' a temperatura ambiente. El paraformaldehído trabaja sobre distancias cortas (2 Å), por lo que prevalecen los contactos entre proteínas y ADN que están interactuando, sobre contactos al azar (Solomon and Varshavsky 1985; Orlando et al. 1997; Jackson 1999).

Terminado este tiempo se transfirió el frasco de cultivo a hielo, se removió el paraformaldehído y se lavó con una solución de glicina 0,125 M en PBS, la cual bloquea al agente entrecruzante. Se adicionó nuevamente glicina para levantar las células con un "scraper" o rastrillo, y se transfirió la suspensión a un tubo Falcon de 15 mL para centrifugarse a 2000g por 5' a 4°C. Mientras se centrifugaban las células, se disgregó con Tripsina (0,25% Tripsina - EDTA) un frasco de cultivo no tratado, para contar células en cámara de Neubauer y estimar el número en los frascos tratados. Después de la centrifugación, se removió el sobrenadante dejando un volumen residual adecuado para llegar a una concentración  $10^6$  células por cada 100  $\mu$ L de glicina 0,125 M, que se alicuotaron en tubos eppendorfs y se guardaron a -80°C hasta la digestión.

## Digestión

Por cada experimento 4C, se utilizaron tres viales de células entrecruzadas, los cuales se prepararon para la digestión. Se descongelaron los tubos en hielo y se centrifugaron a 2000g por 5' a 4°C. Se removió el sobrenadante, y se resuspendió el "pellet" en el buffer de Mbo I, la enzima de restricción de corte frecuente ( $\text{GATC}$ ) elegida para digerir la cromatina. Las enzimas de corte frecuente generan fragmentos de  $\sim 250$  pb, lo cual aumenta la resolución de la metodología al definir mejor el sitio de interacción (Simonis et al. 2007).

Se centrifugó nuevamente en las mismas condiciones y se resuspendió el pellet en un buffer de digestión (compuesto por buffer de Mbo I e inhibidores de proteasas). A la suspensión se le agregó SDS (concentración final 0,3%) y se incubó con agitación a 1400 rpm toda la noche a 37°C, lo que facilitó la apertura de la cromatina para la digestión. Al día siguiente se agregó más buffer de digestión en cuatro pasos secuenciales, con pipeteo suave para evitar la agregación de células o núcleos por la rápida dilución del SDS. Se adicionó Tritón X-100 (concentración final 1,8%) para secuestrar el SDS, y se incubó a 37°C con agitación a 1400 rpm por 1,5 h. Se guardó 50  $\mu$ L de cada uno de los viales, como controles no digeridos para determinar la eficiencia de la digestión.

Para la digestión propiamente dicha se incubaron las células con 300 U de Mbo I por 4h a 37°C con agitación a 1400 rpm; al cabo de este tiempo se adicionó 300 U más para las siguientes 4 hs y finalmente, 200 U por toda la noche a 37°C con agitación. Siempre se repuso el agua evaporada para no afectar la eficiencia de la reacción. Al día siguiente, se tomó 50 µL de cada vial como controles digeridos, y se procedió a determinar la eficiencia de la digestión. El resto de cada vial se mantuvo a 4°C.

### Eficiencia de la digestión

Para determinar si la carnada se digirió apropiadamente y que no afectara la interpretación de los datos finales, se determinó la eficiencia de la digestión de las alícuotas de ADN tomadas a partir de cada vial antes y después de la digestión.

Cada una de estas alícuotas se incubó con SDS (concentración final 1,7%) a 65°C por 20', para terminar la lisis celular y/o la digestión. Se centrifugó a 12000g por 3' y se transfirió el sobrenadante a otro tubo. Se resuspendió el pellet del tubo original y se diluyó el sobrenadante con 83 µL y 127 µL de buffer TE, respectivamente. Las fracciones se trataron con RNAsa (0,3 mg/mL final) y se incubaron a 37°C por 30'. Luego se revirtió el entrecruzamiento en presencia de proteinasa K (0,5 mg/mL final) y NaCl (0,2 M final), y en el caso de la fracción del pellet también con SDS (0,5% final), por toda la noche a 65°C con agitación. Se purificaron las fracciones con Qiagen PCR Purification Kit (Qiagen).

Primero se compararon visualmente las fracciones del sobrenadante de las alícuotas digeridas y no digeridas, en un gel de agarosa 1,2% (tinción con GoodView, SBS Genetech). Después se analizó cuantitativamente la digestión mediante un qPCR que compara la concentración de amplicones de ADN digerido y no digerido con cebadores que flanqueen y no flanqueen sitios de restricción de Mbo I (iQ SYBR Green Supermix, BIORAD).

En el qPCR, la curva de amplificación presenta tres fases: a) temprana, donde la señal de fluorescencia es muy débil, b) exponencial, donde la señal crece exponencialmente, y c) plató, donde los reactivos se terminaron y no se observa incremento de señal. El usuario puede definir un nivel umbral de fluorescencia, por debajo del cual se considera "ruido" de la amplificación. Durante la fase exponencial, el ciclo en el cual la fluorescencia supera el umbral, se conoce como Ct ("cycle threshold"). Para calcular la eficiencia de la digestión se utilizó la siguiente fórmula:

$$\% \text{ eficiencia} = 100 - \frac{100}{2^{(\Delta\text{Ct ADN digerido} - \Delta\text{Ct ADN no digerido)}}$$

Donde  $\Delta Ct$  corresponde a la diferencia entre el Ct del producto de PCR de ADN (digerido o no) amplificado con cebadores que flanquearan al sitio de restricción Mbo I y el Ct del producto amplificado con cebadores que no lo flanquearan (Wang et al. 2014a). Se recomienda que el porcentaje de digestión sea mayor o igual a 85% (Göndör et al. 2008).

### Terminación de la digestión

Después de establecer que la digestión fue suficiente, se continuó con los viales digeridos guardados a 4°C: se detuvo la digestión y se lisaron las células al incubarse en presencia de 1,6% SDS con agitación a 65°C por 20'. Se centrifugaron las células a 16000g por 3' a temperatura ambiente y se transfirió el sobrenadante de cada muestra a un nuevo tubo eppendorf.

En este punto se definió el destino de cada uno de los tres viales: uno se guardó a 4°C (muestra 4C), mientras que a los otros dos se les revirtió el entrecruzamiento con NaCl (0,2 M final), EDTA (1 mM final) y proteinasa K (0,2 mg/mL final) a 65°C toda la noche (controles). Se purificaron con Qiagen PCR Purification Kit (Qiagen) y se cuantificaron con Nanodrop®.

### Ligación

Para la ligación apropiada de la muestra 4C, era necesario que la concentración de los fragmentos de restricción fuese lo suficientemente baja para favorecer la reacción intramolecular de aquellos fragmentos unidos por el mismo complejo ADN–proteínas (0,8 ng/μL) (Göndör et al. 2008). En base a la concentración de los controles (ya no entrecruzados), se estimó la de la muestra 4C, todavía entrecruzada. Uno de los controles no se ligó (control no ligado) y se guardó a -20°C hasta el paso de enriquecimiento por PCR. El otro control se ligó (control ligado) junto con la muestra 4C. Se agregó buffer de ligación (Triton X-100 1%, inhibidor de proteasas, 10 mM Tris-HCl pH 7,8 y BSA 0,1 mg/mL) a 4 μg de ADN, y se incubó por 1,5 h a 37°C. Se agregó T4 DNA ligasa (6000 U) y su buffer, y se incubó por 72 h a 16°C. El volumen de reacción es de 6 mL, para que la concentración final del ADN sea baja. Al cabo de este tiempo, se incubó la mezcla de ligación 1 h a temperatura ambiente, para que la ligasa complete cualquier espacio en el ADN circular.

### Reversión del entrecruzamiento

La reversión del entrecruzamiento de la muestra 4C se llevó a cabo al incubarla a 65°C por toda la noche, en presencia de NaCl (0,2 M final), EDTA (1 mM final) y proteinasa K (0,2 mg/mL final). Los fragmentos de la muestra 4C que estaban unidos a proteínas en común, formaron círculos quiméricos que contenían a la carnada y a los segmentos de ADN que estaban interactuando con ella, constituyendo una biblioteca 4C de los contactos existentes. No fue necesario incluir a los controles ligado y no ligado pues ya se les había revertido el entrecruzamiento antes de la ligación.

### Purificación del ADN total

Se transfirió la muestra 4C (ligada y con reversión del entrecruzamiento) y el control ligado (con reversión y ligado) a un tubo Falcon de 50 mL, donde se precipitó el ADN a -20°C por un período no menor a 12 hs, por el agregado de glicógeno (0,15 mg/mL final), NaAc pH 5,0 (0,3 M final) y 2 volúmenes de etanol absoluto. Se centrifugó a máxima velocidad 30' a 4°C y se lavó el pellet con etanol 70%. Se centrifugó a máxima velocidad 15' a 4°C, y se resuspendió el pellet en 100 uL de H<sub>2</sub>O libre de DNAsa. El ADN se purificó con el Westburg PCR Purification Kit (Invitex) seguido por el Qiagen PCR Purification Kit (Qiagen), eluyendo en 50 µL de buffer TE 0,5x. Se midió la concentración de ADN con Nanodrop® y se diluyó a una concentración final de 20 ng/µL con buffer TE 0,5x. Se guardó la muestra 4C y el control ligado a -20°C.

### Purificación del ADN circular

Debido a su inestabilidad, el ADN circular se purificó inmediatamente antes del enriquecimiento por PCR. Se digirió el ADN lineal de la muestra 4C y del control ligado con exonucleasas I y III, y se incubó a 37°C por 45' seguido de 80°C por 25', para inactivar las enzimas. Se purificó el ADN circular con Westburg PCR Purification Kit (Invitex), donde se eluyó con buffer TE 0,3x conteniendo 10 ng de ADN de *E.coli* sonificado a un tamaño de 300 pb, para prevenir que el ADN circular se adhiriera a las paredes de la columna del kit. Se incubó en la columna toda la noche a temperatura ambiente, y al día siguiente se centrifugó a 16000g por 2'.

## Enriquecimiento por PCR anidado e invertido

Una vez generada la biblioteca 4C, se enriquecen los círculos que contengan a la carnada (potenciador E o promotor de c-Myc) mediante la amplificación con PCR invertido y anidado con cebadores específicos para la carnada, para asegurarse que todos los amplicones se amplifiquen con la misma eficiencia (Göndör et al. 2008). El diseño de los cebadores empleados se detalla más adelante en Materiales y Métodos.

Para el primer PCR, se hicieron tres réplicas por cada muestra (muestra 4C, controles ligado y no ligado). Cada reacción contenía 10  $\mu$ L de ADN circular purificado, Buffer 1x, 0,65  $\mu$ M de cada uno de los cebadores externos (FwdOut y RevOut, ver más adelante en Figura 10 de la sección Diseño de cebadores), 0,35 mM dNTPs, 0,2 mM  $MgCl_2$ , 0,5 ng/ $\mu$ L ADN E.coli sonificado (para la procesividad enzimática), 3,3% DMSO y 0,03 U/ $\mu$ L de enzima.

Las condiciones de ciclado fueron: desnaturalización inicial a 98°C por 30" a 98°C, 4 ciclos extensos (desnaturalización 98°C por 10", annealing por 40', extensión 72°C por 5'), seguido de 30 ciclos más cortos (desnaturalización 98°C por 7", annealing por 2' 30", extensión 72°C por 5'), y una extensión final de 72°C por 7'.

Como molde del segundo PCR, se utilizó una dilución 1/100 en buffer TE del producto del primero, también por triplicado para cada dilución. Cada reacción de este segundo PCR contenía 2,5  $\mu$ L de la dilución, Buffer 1x, 0,35 mM dNTPs, 0,65  $\mu$ M de cada cebador interno (FwdIn y RevIn, Figura 10), 0,15 mM de  $MgCl_2$  y 0,05 U/ $\mu$ L polimerasa. Las condiciones de ciclado fueron: desnaturalización inicial de 94°C por 4', 27 ciclos (desnaturalización 94°C por 30", annealing por 1', extensión 72°C por 4'), y una extensión final de 72°C por 7'. Los 9 productos de PCR finales de cada muestra (tres del primer PCR y sus triplicados en el segundo), se purificaron con Qiagen PCR Purification Kit.

En el control ligado se espera que amplifique una banda del tamaño esperado al del diseño del PCR, que corresponde al evento de autoligación del fragmento de restricción que contiene a la carnada. En la muestra 4C se debe observar un "smear" que representa todos los posibles interactores de la carnada, y en el control ligado no se debe amplificar nada, pues no se ligó (Figura 5).

## **Secuenciación de bibliotecas 4C**

### Selección de amplicones 4C

Se secuenció solamente la biblioteca de la muestra 4C y no los controles ligado y no ligado, porque éstos sólo sirvieron a los efectos de corroborar la digestión y ligación de la cromatina.

Debido a que los amplicones del tamaño esperado corresponderían a eventos de autoligación de la cromatina, y que serían los más abundantes en la biblioteca de la muestra 4C, se secuenciaron los amplicones de mayor tamaño, conteniendo como inserto a las regiones interactoras. Esto se logró a través una electroforesis en gel de agarosa 1,2% de cada una de las bibliotecas 4C (de las muestras 4C del Potenciador E y del promotor de c-Myc), extracción con QIAquick Gel Extraction Kit (QIAGEN), y cuantificación con el fluorímetro Qubit® dsDNA HS Assay (Life Technologies).

### Optimización de la preparación de las librerías de secuenciación

Los productos de PCR fueron secuenciados con Ion Torrent (Life Technologies). Cada librería de secuenciación (Potenciador E y promotor de c-Myc) se preparó según el protocolo de Ion Xpress™ Plus Fragment Library Kit, donde 100 ng de los productos de PCR se fragmentaron de forma enzimática para una distribución de tamaño de 400 pb. El tiempo de reacción promedio para obtener dicha distribución es de 8'; pero como el protocolo está optimizado para ADN genómico y puede variar según el contenido GC de la muestra, se probó con distintos tiempos dentro de un rango entre 5 y 12' sugerido por el fabricante. Esto se testeó con la fragmentación a 5, 7 y 9' de los amplicones del potenciador E y se analizó con Agilent Bioanalyzer 2100 System con el kit Agilent High Sensitivity DNA Analysis (Agilent). Una vez optimizada la reacción, se continuó con la ligación de los adaptadores con código de barra para identificar a cada librería al final de la secuenciación.

### Amplificación clonal por PCR en emulsión y secuenciación

Los productos intermedios de la fragmentación y ligación se purificaron con perlas magnéticas del sistema de purificación de PCR Agencourt® AMPure® XP (Beckman Coulter). Se confirmó la distribución de cada librería (ahora con adaptadores) y se cuantificó en un equipo Agilent Bioanalyzer 2100 System con el kit Agilent High Sensitivity DNA Analysis.

El PCR en emulsión se llevó a cabo en el equipo Ion OneTouch™ 2 System de acuerdo al protocolo del kit Ion PGM™ Template OT2 400 Kit, mezclando ambas librerías en una sola reacción. La secuenciación se corrió en un Ion 318™ Chip en el equipo Ion PGM™ System, según el protocolo del kit Ion PGM™ Sequencing 400 Kit v2 (Figura 6).

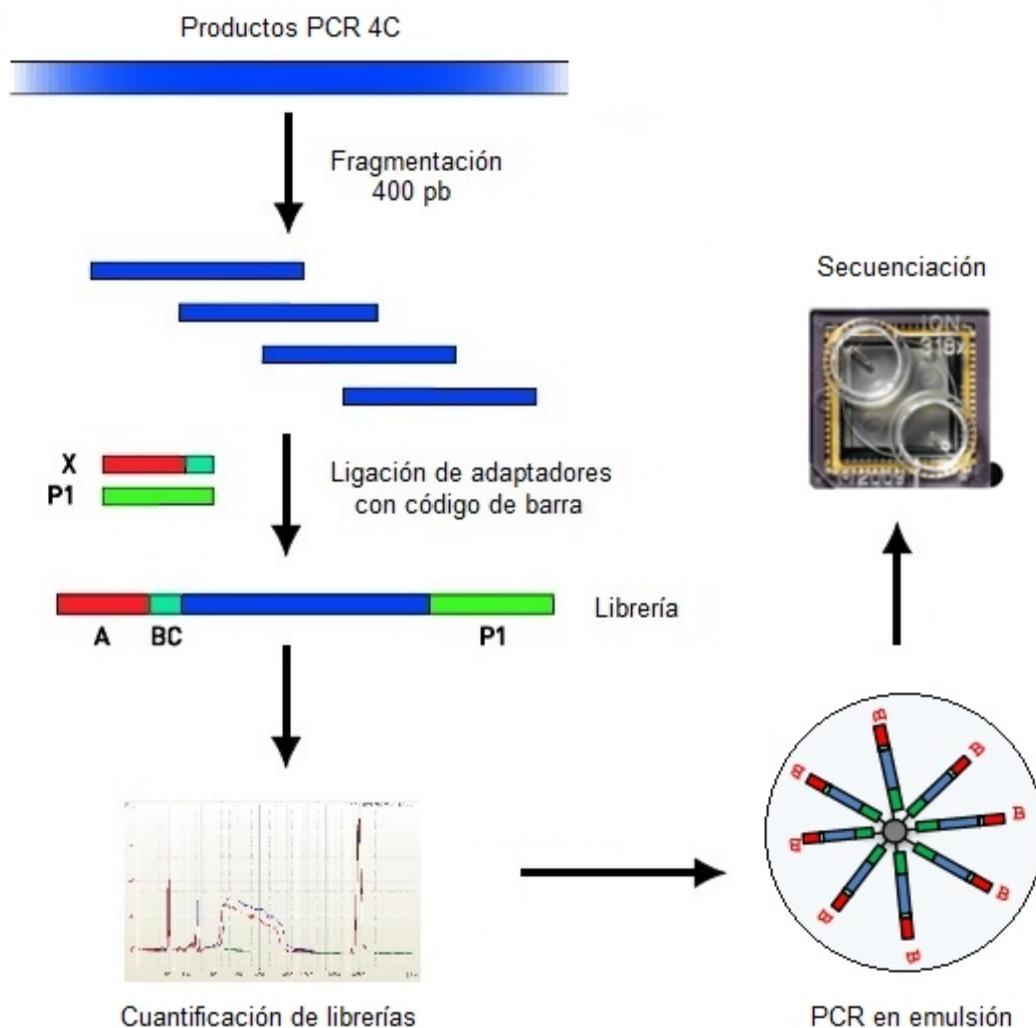


Figura 6 – Secuenciación en Ion Torrent: preparación de las librerías por fragmentación y ligación de adaptadores con código de barra, cuantificación con Bioanalyzer, amplificación clonal por PCR en emulsión y secuenciación (modificado de material de Life Technologies).

### Análisis de datos de secuenciación

Se generó un archivo .fastq para cada librería con información de las lecturas (nombre, secuencia, calidad), con el complemento FastQCreator del programa Torrent Suite asociado al secuenciador. También se hizo un análisis de calidad de las lecturas con FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Una vez obtenidos los archivos .fastq de cada librería, se utilizó el programa Sickle para remover los nucleótidos con un puntaje de calidad inferior a 20<sup>1</sup>, y se conservaron sólo las lecturas con longitud mayor o igual a 50 nucleótidos (sickle se -f -q 20 -t sanger -l 50) (Joshi y Fass 2011, disponible en <https://github.com/najoshi/sickle>) (Figura 7).

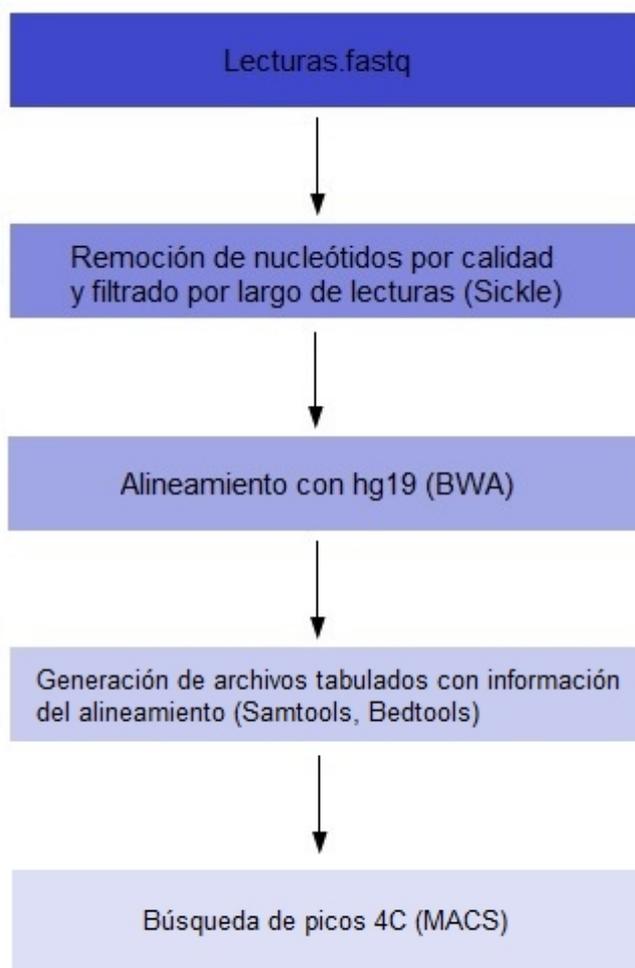


Figura 7 – Procesamiento de los datos 4C.

Las lecturas procesadas se alinearon contra la secuencia de referencia del genoma humano (hg19 o GRCh37), la cual se obtuvo desde el Genome Browser de la Universidad de California, Santa Cruz (UCSC) (<http://hgdownload.cse.ucsc.edu>). Para el alineamiento se utilizó el programa BWA (Burrows-Wheeler Aligner) (Li and Durbin 2009), que se especializa en alinear secuencias contra un gran genoma de referencia como lo es el humano (Figura 7). Como el rango del largo de las lecturas resultó ser amplio, se empleó el algoritmo MEM de BWA (Maximal Exact Match), que además reporta lecturas quiméricas, es decir, lecturas cuyas fracciones alinean en dos sitios genómicos, tales como suele suceder en translocaciones.

---

<sup>1</sup> En la escala de Phred para Sanger, 20 corresponde a una tasa de error de 1 nucleótido en cada 100.

El archivo de salida `.sam` (Sequence Alignment/Map) generado en el alineamiento se convirtió en su versión binaria `.bam` (Binary Alignment/Map), mediante Samtools (Li et al. 2009). Este programa indexa el genoma de referencia (`samtools faidx`), hace la conversión propiamente dicha (`samtools view -bS`), y finalmente ordena e indexa el archivo `.bam` por las coordenadas cromosómicas (`samtools sort`; `samtools index`) (Figura 7). Con Samtools (`samtools view`) y comando de linux (`grep`, `awk`, `sed`, `sort`, `uniq` y `wc`) se contabilizaron las lecturas alineadas, así como las quiméricas (con fracciones correspondientes a dos sitios genómicos distintos) y no quiméricas (toda la lectura corresponde a un solo sitio).

### **Búsqueda de picos 4C**

Para buscar regiones genómicas con una mayor densidad de lecturas, definidas como picos (Figura 7), se utilizó a MACS (Model-based Analysis of ChIP-Seq) (Zhang et al. 2008), un algoritmo implementado en Python usualmente empleado para reportar picos en estudios de ChIP-Seq (combinación de ChIP con secuenciación masiva). MACS calcula localmente el promedio y la varianza de las lecturas distribuidas en el genoma “background”, para definir los picos. Una ventaja de MACS para el estudio de datos 4C es que no es necesario utilizar un control como se hace en ChIP.

MACS requiere la conversión del archivo `.bam` ('alineamiento.bam') proveniente del alineamiento, en `.bed` ('alineamiento.bed'), un archivo tabulado que contiene el nombre del cromosoma donde alineó cada secuencia, así como la posición de inicio y de final de las mismas. La conversión de `.bam` a `.bed` se hizo con Bedtools (`bedtools bamtoBED -i`) (Quinlan and Hall 2010), para finalmente identificar los picos con MACS (`macs14 -t -f BED -n -g -w`), donde los parámetros `g` y `w` corresponde al tamaño del genoma de referencia ( $2.7e^9$ ).

MACS generó otro archivo `.bed` conteniendo localización genómica de los picos 4C, un `.xls` con mayor información que `.bed` (extensión de los picos, número de lecturas en cada uno, p-valor, etc.) ('picos.xls'), y un `.wig` que muestra a los picos 4C como una distribución a lo largo del genoma ('picos.wig'). Con comandos de linux se obtuvo el número de picos 4C identificados (`wc -l`), la extensión de los mismos contenida en la columna 4 del archivo `.xls` (`awk '{print$4}'`), los picos con mayor número de lecturas (`sort -nk6, tail`) y los picos en común a los datos 4C del potenciador E y c-Myc (`bedtools intersect -a -b -wa, wc -l`).

## **Análisis de datos 4C**

### Genes cercanos a picos 4C en común y a los más abundantes

Como una primera aproximación al estudio de los picos 4C obtenidos, se buscaron los genes más cercanos a los picos en común del potenciador E y c-Myc, y a los picos más abundantes para cada carnada, mediante la herramienta GREAT (Genomic Region Enrichment of Annotations Tool) disponible en el UCSC Genome Browser. GREAT examina las regiones de los picos e indica la distancia respecto al sitio de inicio de la transcripción o TSS (Transcription Start Site) de los genes más cercanos. También se estudió la participación de los genes cercanos en enfermedades y desórdenes así como también sus funciones celulares y moleculares a través de Ingenuity Pathway Analysis (IPA<sup>®</sup>, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)).

### Distribución en el genoma y estructuras génicas

Para estudiar la distribución de todos los picos 4C en el genoma y su potencial asociación con estructuras génicas (promotores, exones, intrones, UTRs, regiones río abajo de genes), se empleó el programa CEAS (Cis-regulatory Element Annotation System) (Shin et al. 2009), también utilizado para CHIP-Seq. CEAS compara el enriquecimiento de los picos 4C respecto al genoma mapeable (<http://liulab.dfci.harvard.edu/CEAS/usermanual.html>) y requiere los archivos 'picos.bed' y 'picos.wig' generados por MACS.

### Interpretación funcional

Para analizar más profundamente los datos 4C e intentar darles un significado funcional, se planteó como punto de partida intersectar los picos 4C con datos de CHIP y DNAsa-Seq, que permitieran identificar regiones promotoras o elementos potenciadores, activos e involucrados en interacciones de largo alcance.

Para complementar los resultados de las intersecciones del potenciador E, también se analizó el nivel de expresión de cada región candidata y se buscó literatura que la relacionara con cáncer.

### Intersección con datos de ChIP-Seq y DNAsa-Seq

Para predecir genes candidatos a ser regulados por el potenciador E, y elementos reguladores del gen c-Myc, se interseccionaron los picos 4C con datos de ChIP y DNAsa-Seq. Para lograr esto se cargaron los archivos 'picos.bed' de cada carnada generados por MACS, y datos de ChIP y DNAsa-Seq al UCSC Genome Browser (Tabla 1). La intersección se realizó mediante la herramienta UCSC Table Browser disponible en UCSC Genome Browser, que generó un “track” por cada cruce de datos.

Tabla 1 – Datos de secuenciación utilizados para interseccionar con datos 4C				
Dato	Línea celular	Fuente	Acceso datos crudos	Descarga .BED
DHS	LNCaP	ENCODE	GSM816634 (GEO <sup>2</sup> )	UCSC*
H3K4me1	LNCaP	Yu et al. 2010	GSM353634 (GEO)	Cistrome
H3K4me3	LNCaP	Yu et al. 2010	GSM353626 (GEO)	Cistrome
H3K27ac	LNCaP	Hazelett et al. 2014	GSM1249447 (GEO)	UCSC*
CTCF	LNCaP	ENCODE	GSM1006874 (GEO)	UCSC*
TCF4	LNCaP	Hazelett et al. 2014	GSM1249449 (GEO)	UCSC*
p300	LNCaP	Wang et al. 2011	GSM686944 (GEO)	Cistrome
RNA Pol II	VCaP	Yu et al. 2010	GSM353623 (GEO)	Cistrome
RAD21	MCF-7	Schmidt et al. 2010	E-TABM-828 (ArrayExpress)	ArrayExpress
STAG1	MCF-7	Schmidt et al. 2010	E-TABM-828 (ArrayExpress)	ArrayExpress

\* Disponibles como track cargado por Hazelett et al. ([http://genome.ucsc.edu/cgi-bin/hgTracks?hgS\\_doOtherUser=submit&hgS\\_otherUserName=hazelett&hgS\\_otherUserSessionName=pca](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=hazelett&hgS_otherUserSessionName=pca))

En el caso del potenciador E, se interseccionaron los picos 4C con datos de ChIP-Seq de H3K4me3, un marcador epigenético de regiones promotoras (Figura 8). El track resultante se ingresó en GREAT y se obtuvieron todos los genes cercanos a los picos 4C que contenían al marcador.

Aquellos genes donde la intersección se localizaba entre -2000 pb y +500 pb respecto a su TSS, se seleccionaron mediante awk. A su vez, esta intersección se cruzó con datos de marcadores de cromatina abierta tales como ChIP-Seq de sitios de unión de RNA Pol II y DNAsa-Seq para sitios de hipersensibilidad a la DNAsa I o DHS (DNase I Hypersensitivity Sites), y ChIP-Seq para sitios de unión de proteínas involucradas en interacciones de largo alcance (CTCF, STAG1, RAD21) y del factor de transcripción TCF4 (Figura 8).

<sup>2</sup> GEO: Gene Expression Omnibus)

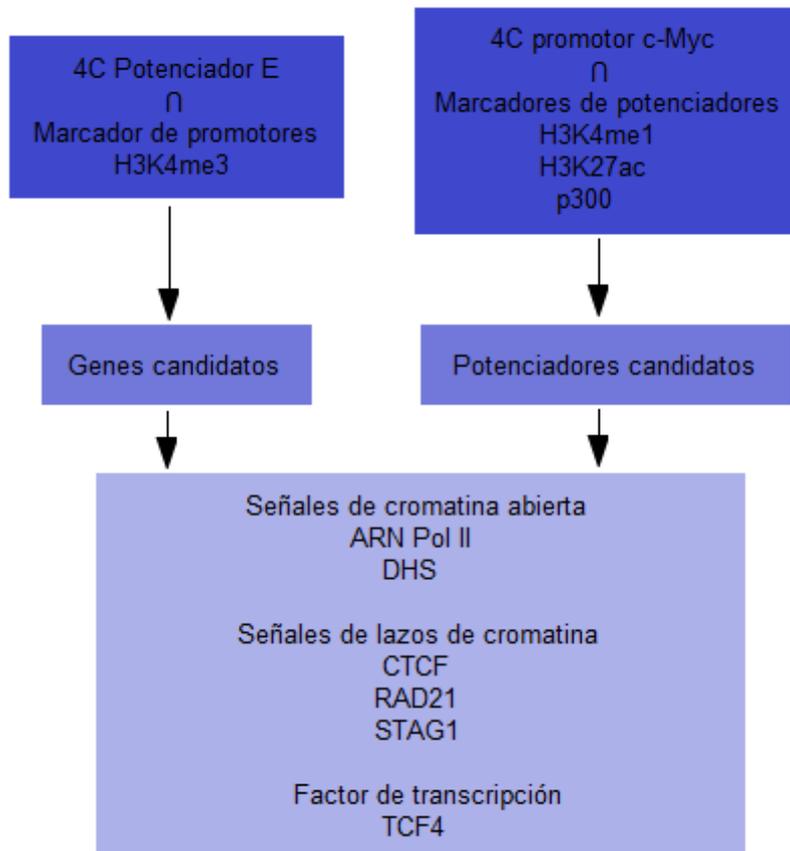


Figura 8 – Interpretación funcional de picos 4C. Los picos de cada carnada se intersecta con marcadores específicos de promotores o potenciadores, según sea el caso. En las regiones candidatas también se estudia la presencia de sitios de unión para otras proteínas de interés (lazos de cromatina, accesibilidad, factores de transcripción).

En el caso de c-Myc, se intersectaron los picos 4C con datos de ChIP-Seq de tres marcadores para elementos potenciadores: las modificaciones de histonas H3K4me1 y H3K27ac, y sitios de unión para p300, una enzima acetiltransferasa de histonas (Figura 8). El track resultante se cruzó con datos ChIP-Seq y DNAsa-Seq de marcadores de cromatina abierta (RNA Pol II, DHS), proteínas involucradas en interacciones de largo alcance (CTCF, STAG1, RAD21) y el factor TCF4 (Figura 8).

### *Expresión en muestras clínicas*

Como estudio complementario, se estudió la expresión de los genes candidatos a ser regulados por el potenciador E, en muestras clínicas de tumores de próstata, a través de la base de datos Oncomine (Rhodes et al. 2007), que colecta, estandariza y analiza datos de microarreglos publicados.

Se llevó a cabo un análisis de Expresión Diferencial en Oncomine, entre carcinoma de próstata y glándula normal de próstata, el cual utiliza el test t de Student para el análisis estadístico (two-sided). Este análisis se realizó sobre 14 estudios de microarreglos de cáncer de próstata, que en total sumaron 261 controles y 713 casos.

#### *Literatura vinculada a cáncer*

También se recurrió a la literatura para obtener información que sustentase, indicase o sugiriese algún vínculo de los genes posiblemente regulados por el potenciador E, con diversos aspectos de cáncer, tales como desarrollo y progresión, expresión específica de tejido, etc.).

#### Visualización de datos e interacciones

Los datos 4C, CHIP y DNAsa-Seq se visualizaron y exploraron a nivel genómico en IGV (Integrative Genomic Viewer) (Thorvaldsdóttir et al. 2013). Los contactos entre cada carnada y sus correspondientes interactores se visualizaron en una configuración circular con el programa Circos (Krzywinski et al. 2009).

#### **Chromosome Conformation Capture o 3C**

Las interacciones genómicas que resultaron ser más interesantes o pertinentes para los objetivos de esta tesis, se confirmaron mediante la metodología Chromosome Conformation Capture o 3C (Dekker et al. 2002), de acuerdo al protocolo utilizado por Hagège et al. (2007). Mientras 4C permite responder con quienes interactúa una región de interés, 3C permite responder si existe contacto físico entre dos regiones de interés conocidas *a priori*. En esta tesis dichas regiones corresponderían a las carnadas 4C (Potenciador E y promotor de c-Myc) con cada uno de sus posibles interactores identificados.

El procedimiento implicó: a) entrecruzamiento de la cromatina para formar complejos ADN – proteínas que estuviesen cercanos en el núcleo; b) digestión enzimática del ADN con una enzima de restricción de corte raro; c) ligación de los fragmentos de restricción unidos por el mismo complejo, en condiciones que generen moléculas lineales quiméricas formadas por las distintas secuencias interactoras; d) reversión del entrecruzamiento; e) purificación del ADN quimérico; y f) confirmación de un contacto en particular por PCR con cebadores específicos para las dos regiones de interés, que flanqueen el sitio de restricción (Figura 3).

## Entrecruzamiento

Cuando las células LNCaP alcanzaron un 80% de confluencia, se les removió el medio de cultivo y se lavaron dos veces con PBS a temperatura ambiente. Se disgregaron al incubar a 37°C por 5' con Tripsina (0,25% Tripsina – EDTA). Se resuspendió el pellet en medio RPMI 1640 y se transfirió la suspensión a un tubo Falcon de 15 mL para centrifugarse a 400g por 1' a temperatura ambiente. Se descartó el sobrenadante y se resuspendió el pellet en un volumen de PBS para llegar a una concentración  $10^7$  células por cada 500  $\mu$ L. En 3C se entrecruzó la cromatina de las células disgregadas (en suspensión, no adherentes), por incubación con paraformaldehído al 2% en PBS (preparado fresco), durante 10' a temperatura ambiente.

Terminado este tiempo se transfirió el tubo Falcon a hielo y se agregó una solución de glicina 1 M en PBS, la cual bloquea al agente entrecruzante. Se centrifugó a 225g por 8' a 4°C, se removió el sobrenadante y se guardaron los pellets a -80°C hasta la digestión.

## Digestión

Se resuspendió cada pellet en el buffer de Bgl II (1,2x), la enzima de restricción elegida para digerir la cromatina (A<sup>GATCT</sup>). La suspensión se transfirió a un tubo eppendorf, se le agregó SDS (concentración final 0,3%) y se incubó a 37°C por 1 h con agitación a 900 rpm. Se adicionó Tritón X-100 (concentración final 2%) para secuestrar el SDS y se incubó a 37°C por 1 h con agitación a 900 rpm. Para la digestión se agregó 400 U de Bgl II y se mantuvo a 37°C por toda la noche con agitación.

## Ligación

Se terminó la digestión por la adición de SDS (concentración final 1,6%) e incubación a 65°C por 20-25' con agitación a 900 rpm. Se transfirió la reacción a un tubo Falcon de 50 mL, donde se agregó buffer de ligación 1,15x y Tritón X-100 (concentración final 1%) para incubarse a 37°C por 1 h con agitación. Se agregó T4 DNA ligasa (100 U) por 4 h a 16°C seguido de 30' a temperatura ambiente, para que la enzima completara cualquier espacio en el ADN ligado.

### Reversión del entrecruzamiento

Se revirtió el entrecruzamiento en presencia de proteinasa K (300 µg final) y se incubó a 65°C por toda la noche.

### Purificación del ADN

Se agregó RNAsa (300 µg final), y al cabo de 30-45' a 37°C, se hizo una extracción con fenol-cloroformo. Se transfirió el sobrenadante a un nuevo tubo Falcon de 50 mL, se precipitó el ADN con 0,1 volúmenes de NaAc 2 M pH 5,6 y 2 volúmenes de etanol 100% a -80°C por 1 h. Después se centrifugó a 2200 g por 45' a 4°C, se removió el sobrenadante y se lavó el pellet con etanol 70%. Se centrifugó a 2200g por 15' a 4°C, se removió el sobrenadante, se secó el pellet y se resuspendió en Tris 10 mM pH 7,5. Las alícuotas se guardaron a -20°C.

### Amplificación de moléculas conteniendo a las regiones que interactúan

Como resultado de las condiciones de la ligación, se obtuvieron moléculas quiméricas lineales que representaban los contactos que existían en la célula. La confirmación de cada contacto identificado previamente por 4C (carnada 4C – interactor) se hizo a través de PCR con cebadores específicos para cada región, y que flanquearan un sitio Bgl II. Se diseñaron los cebadores de acuerdo a Naumova et al. (2012) con el programa CLC Genomics Workbench v6.5.

Cada reacción de 20 µL contenía Buffer 1x, 0,65 µM de cada uno de los cebadores, 0,35 mM dNTPs, 0,2 mM MgCl<sub>2</sub>, 0,03 U/µL de enzima y 20 ng de ADN. Las condiciones de ciclado fueron: desnaturalización inicial a 98°C por 30" a 98 °C, 4 ciclos (desnaturalización 98°C por 10", annealing por 40', extensión 72°C por 5'), y una extensión final de 72°C por 7'.

### **Diseño de cebadores**

Todos los cebadores utilizados en esta tesis se diseñaron con el programa CLC Genomics Workbench v6.5. Para la eficiencia de la digestión, se diseñaron dos juegos de cebadores que flanquearan sitios de corte de la enzima Mbo I en la región promotora de c-Myc (F1-R1 y F2-R2), y otro juego que no lo hiciera F3-R3 (Figura 9), donde R2 y R3 poseen la misma secuencia.

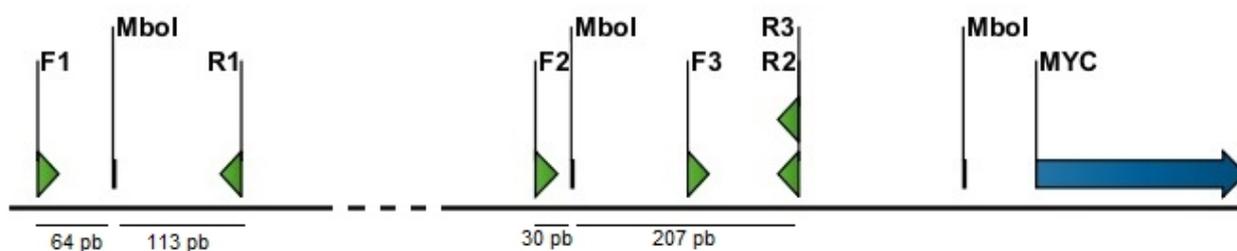


Figura 9 – Cebadores para determinar la eficiencia de la digestión en la región promotora de c-Myc. Los triángulos verdes representan a los cebadores, y las barras verticales a los sitios de restricción de Mbo I. Debajo se indica la distancia de cada cebador al sitio Mbo I correspondiente.

En la Tabla 2 se muestra la secuencia, el largo y la temperatura de “melting” de cada cebador, así como el E-valor de su alineamiento con el genoma de referencia hg19 a través de BLAST (Basic Local Alignment Search Tool), herramienta disponible en el National Center for Biotechnology Information o NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). También se muestra el tamaño esperado del amplicón producido por cada par de cebadores.

Tabla 2 – Cebadores utilizados para medir la eficiencia de la digestión					
Nombre	Secuencia 5' – 3'	Largo (nt)	Tm (°C)	Amplicón (pb)	E-valor
F1	AATTAGCTCAAGACTCCCC	20	58	178	0.015
R1	CCGTAACTCTTTCCTCCC	19	58		0.041
F2	CCTTCTTTCCTCCACTCTCCC	21	61	238	0.005
R2	TCCCTCTCAAACCCTCTCCC	20	61		0.015
F3	CCGCCTGCGATGATTATAC	20	60	98	0.015
R3	TCCCTCTCAAACCCTCTCCC	20	61		0.015

Para el enriquecimiento por PCR invertido y anidado, se analizaron los sitios de restricción de Mbo I en la secuencia de referencia del genoma humano (hg19) y se diseñaron las moléculas circulares y los cebadores específicos para cada carnada. Para el diseño de los cebadores se consideró que la molécula circular contenía sólo la secuencia de la carnada, sin inserto o interactor, por lo que el tamaño esperado del amplicón correspondía al evento de autoligación de la molécula (Figura 10).

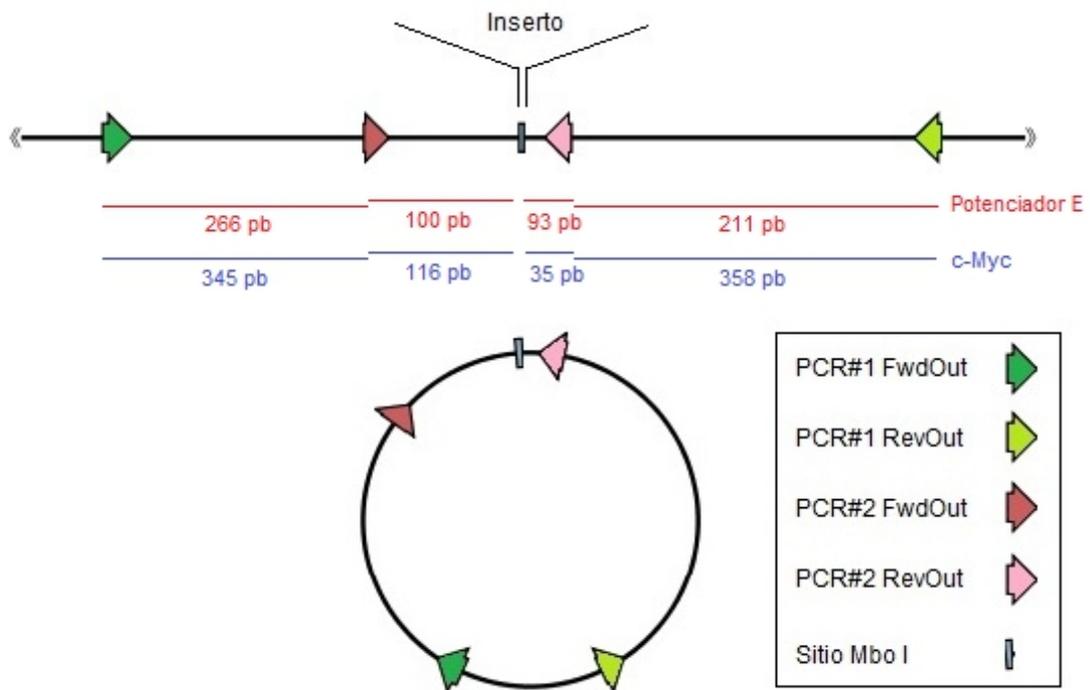


Figura 10 – Cebadores para el enriquecimiento 4C. Abajo se representa la molécula circular que contiene sólo la secuencia de la carnada, donde la barra vertical celeste corresponde al sitio de restricción de Mbo I donde se inserta el interactor, y los triángulos corresponden a los cebadores. Arriba se muestra una linealización de la molécula circular para mejor visualización de la posición de los cebadores, y las distancias de cada cebador de cada carnada al sitio Mbo I.

En la Tabla 3 se muestra la secuencia, el largo y la temperatura de “melting” de cada cebador para cada carnada, así como el E-valor de su alineamiento con el genoma de referencia hg19 a través de BLAST. También se muestra el tamaño esperado del amplicón producido por cada par de cebadores.

Tabla 3 – Cebadores utilizados para el enriquecimiento durante 4C						
	Nombre	Secuencia 5' – 3'	Largo (nt)	Tm (°C)	Amplicón (pb)	E-valor
Potenciador E	PCR#1 FwdOut	GGTATGAATAGGAGAGCT	18	53	193	0.16
	PCR#1 RevOut	GGGAAAGTTGATGGTGAT	18	54		0.16
	PCR#2 FwdIn	GAAGCTTAGAAGGGCATT	20	54		0.015
	PCR#2 RevIn	AAGAGGGAGGTATCAACAG	19	54		0.041
Promotor c-Myc	PCR#1 FwdOut	TAAACAGACGCCTCCCGCAC	20	63	151	0.015
	PCR#1 RevOut	CCCAATTTCTCAGCCAGGTTTC	22	62		0.001
	PCR#2 FwdIn	TACCCTTCTTTCTCCACTC	20	56		0.015
	PCR#2 RevIn	CGTAACTCTTTCTCCCC	19	56		0.041

Para confirmar las interacciones 4C a través de 3C-PCR, se diseñó un cebador específico para cada uno de los fragmentos de restricción generados por Bgl II que contuvieran a las regiones identificadas por 4C (Dekker et al. 2012). Cada cebador se localizaba sobre uno de los extremos de cada fragmento, apuntando hacia el sitio de restricción (Figura 11A). Cuando dichos extremos se ligan entre sí, se amplifica el sitio de restricción/ligación junto con las secuencias río abajo de cada cebador (Figura 11B y C).

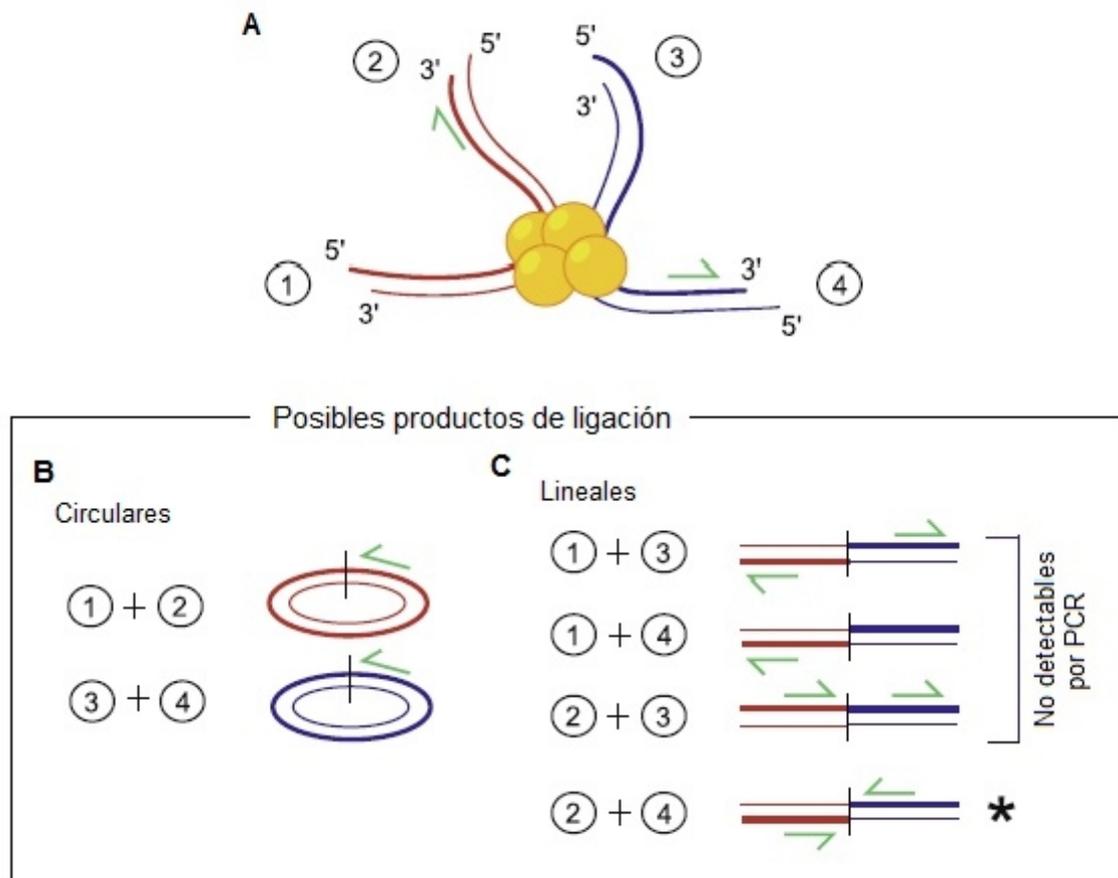


Figura 11 – Cebadores para validación por 3C-PCR. Arriba. A. Situación de entrecruzamiento simplificado, donde dos fragmentos de restricción (uno de color rojo, el otro azul) están unidos por un mismo complejo. Los extremos de cada fragmento están numerados del 1 al 4, y los extremos de cada hebra están indicados (5', 3'). En color verde se indica la posición de cada cebador en un solo sentido en uno solo de los extremos de cada fragmento. Abajo. Posibles productos de ligación. B. Se pueden formar círculos por la autoligación de los fragmentos de restricción, que no amplificarán. C. Se pueden formar moléculas lineales quiméricas, donde solamente una será detectada por PCR, cuando los extremos sobre los cuales se diseñaron los cebadores se ligan entre sí (Modificado de Naumova et al. 2012).

En la Tabla 4 se muestra la secuencia, el largo y la temperatura de “melting” de cada par de cebadores para 3C-PCR, así como el E-valor de su alineamiento con el genoma de referencia hg19 a través de BLAST y el tamaño esperado del amplicón.

Tabla 4 – Cebadores utilizados para 3C-PCR					
Par Interactor	Secuencia 5'-3'	Largo (nt)	Tm (°C)	Amplicón (pb)	E-valor
Potenciador_1	ATAGAGAGAGATGCTGATTGGC	22	59	227	8,00E-004
ACBD3	ATATGATGAGTTTTGGCACAGG	22	59		8,00E-004
Potenciador_1	ATAGAGAGAGATGCTGATTGGC	22	59	267	8,00E-004
BCL9	GAGTATGGGAAAAGCTGAAGGAAAAG	25	61		2,00E-005
Potenciador_1	ATAGAGAGAGATGCTGATTGGC	22	59	269	8,00E-004
HPSE	AGAGGATGACAGAGAAAAGCAGGAAG	25	63		2,00E-005
Potenciador_2	AAACCATAGAGAGAGATGCTGA	22	58	383	8,00E-004
RTF1	GACTGGCAATTAGAAGGAGTAG	22	57		8,00E-004
Potenciador_2	AAACCATAGAGAGAGATGCTGA	22	58	375	8,00E-004
SMARCC1	TTTTGAAAGTATATTGGCCGGG	22	57		8,00E-004
c-Myc_1	TCCTCGGATTCTCTGCTCTCCTC	23	63	351	3,00E-004
15q15.3	CGTCTCTCCAATCCTTTCTGCT	23	62		8,00E-004
c-Myc_1	TCCTCGGATTCTCTGCTCTCCTC	23	63	358	3,00E-004
19p13.3	CAACACAGCGAGACAATTTTTTC	22	58		8,00E-004
c-Myc_1	TCCTCGGATTCTCTGCTCTCCTC	23	63	292	3,00E-004
1q25.3	TATCTAAAGTTCTCTCCTTCCTGCC	25	60	489	2,00E-005
c-Myc_3	CTTCCCCTACCCTCTCAACGAC	22	64		8,00E-004
21q22.3	ATCTTGGCTCACTGCAACCTCC	22	64		8,00E-004
c-Myc_2	GTCAAAAGTGGGCGGCTGGAT	21	65	322	3,00E-003
9p23	ACCAGGAAAGATCAAGACCCAAG	23	62		3,00E-004
c-Myc_2	GTCAAAAGTGGGCGGCTGGAT	21	65	191	3,00E-003
1p34.4	TGATAGGAGGAGAAGCAGGAGTGG	24	64		3,00E-004
c-Myc_1	TCCTCGGATTCTCTGCTCTCCTC	23	63	200	3,00E-004
1p11.2	TGACTTTTCTCCCGTATCCAGC	22	60		8,00E-004
c-Myc_3	CTTCCCCTACCCTCTCAACGAC	22	64	432	8,00E-004
2q21.2	GGTCTGATGTGAAGTTGTCTT	22	58		8,00E-004

### Comparación con otros datos de interacciones en la línea LNCaP

Para corroborar los interactores del potenciador E identificados en esta tesis mediante 4C, se comparó con los publicados por Wang et al. (2014a), quienes estudiaron varias líneas celulares derivadas de cáncer de próstata, incluida LNCaP. Este grupo desarrolló una metodología derivada de 3C conocida como 3C-MTS (3C–Multiple Target Sequencing), en la cual se prepara una librería de secuenciación a partir de la biblioteca 3C y se multi-capturan las moléculas que contengan fragmentos pertenecientes a la región 8q24 (incluyendo al potenciador E) con sondas específicas marcadas con biotina. Posteriormente los fragmentos seleccionados son amplificados y secuenciados (Wang et al. 2014a).

## VI. RESULTADOS

### Generación de bibliotecas 4C

Las bibliotecas 4C para cada región de interés, potenciador E y promotor de c-Myc, se obtuvieron siguiendo el protocolo utilizado por Göndör et al. (2008).

#### Eficiencia de digestión

Durante la preparación de las bibliotecas 4C, se determinó la eficiencia de la digestión de la cromatina de manera cualitativa a través de una electroforesis en gel de agarosa, y de manera cuantitativa por qPCR. En la electroforesis se corrieron las fracciones del sobrenadante de las alícuotas que se extrajeron antes y después de la digestión con Mbo I, y a las que se les revirtió el entrecruzamiento. Las tres fracciones habrían alcanzado el 85% de digestión, al comparar con los resultados obtenidos por Göndör y colaboradores (Figura 12). La fracción con mayor concentración de ADN (carril 5, Figura 12) fue elegida posteriormente como muestra 4C.

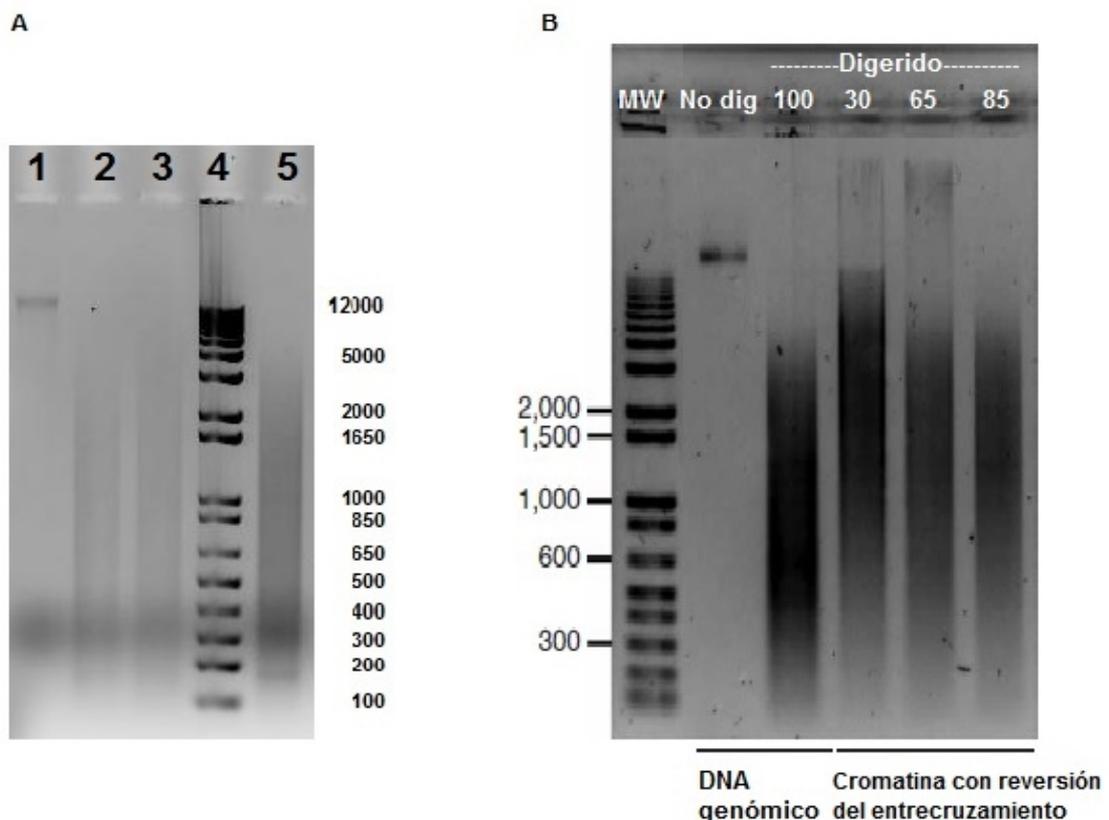


Figura 12 – Eficiencia de la digestión de la cromatina. A) ADN digerido y no digerido con Mbo I durante el experimento 4C de la presente tesis. Carriles: 1 – ADN no digerido; 2, 3 y 5: ADN digerido de distintas alícuotas 4C; 4 – Ladder. B) Figura modificada de Göndör et al. (2008), a modo de comparación.

Luego de corroborar visualmente la eficiencia de la digestión a través de la electroforesis, se determinó cuantitativamente el porcentaje de ADN digerido a través de un qPCR con cebadores que flanquearan sitios Mbo I y que no lo hicieran, en ADN digerido y no digerido (Tabla 2). Como no se logró amplificar el ADN no digerido con los cebadores F1 – R1, se calculó la eficiencia de la digestión con los otros cebadores. Se utilizó la fórmula presentada en Materiales y Métodos para determinar la eficiencia de la digestión, la cual alcanzó el 78%.

#### Enriquecimiento en moléculas conteniendo la carnada

El enriquecimiento de cada biblioteca 4C en las moléculas circulares que contenían a la carnada correspondiente, se hizo a través de PCR invertido y anidado con cebadores específicos para la carnada (Tabla 3). Para el diseño de los cebadores se consideró como molde al fragmento de restricción que contenía a la carnada en una configuración circular, es decir, como si dicho fragmento se ligara sobre sí mismo (Figura 10).

En el control no ligado, al cual se le revirtió el entrecruzamiento pero no se ligó, no se formaron moléculas circulares y por lo tanto no amplificó (Figura 13). En la muestra 4C, la cual se ligó antes de revertir el entrecruzamiento, se formaron moléculas quiméricas (carnada e interactores), por lo que se observó un amplicón correspondiente al evento de autoligación de la carnada, con un tamaño esperado de 150 pb para c-Myc y 193 pb para el potenciador E (Figura 13).



Figura 13 – Enriquecimiento de bibliotecas 4C por PCR invertido y anidado. Izquierda: Myc como carnada; carriles: 1 – Control no ligado; 2 – Control ligado; 3 – Muestra 4C; 4 – Ladder. Derecha: Muestra 4C del potenciador E como carnada.

También se observó un smear de mayor tamaño que llegaba hasta 1500 pb aproximadamente, con algunas bandas pronunciadas, cuyo conjunto representa la amplificación de las distintas secuencias interactoras insertas en el fragmento de la carnada (Figura 13). En el caso de c-Myc, las bandas pronunciadas tenían un tamaño cercano a 350 y 550 pb, mientras que para el potenciador E, los tamaños eran de 300, 400, 650 y 700 pb (Figura 13).

En el control ligado, al cual se le revirtió el entrecruzamiento previo a la ligación, se esperaba que sólo amplificara un producto del tamaño esperado (150 pb para c-Myc, 193 pb para el potenciador E). Sin embargo, se observó además un smear similar al de la muestra 4C pero de menor intensidad (Figura 13).

## Secuenciación de bibliotecas 4C

### Optimización del tiempo de fragmentación

Debido a que el smear del PCR anidado e invertido iba de 150-200 hasta 1500 pb (Figura 13), se fragmentaron los amplicones para una distribución de tamaño de 400 pb, el largo que se podía secuenciar. Para optimizar la fragmentación se incubaron los productos de PCR del potenciador E a 5, 7 y 9' de reacción y se analizaron en Agilent Bioanalyzer 2100 System con el kit Agilent High Sensitivity DNA Analysis (Agilent) (Figura 14).

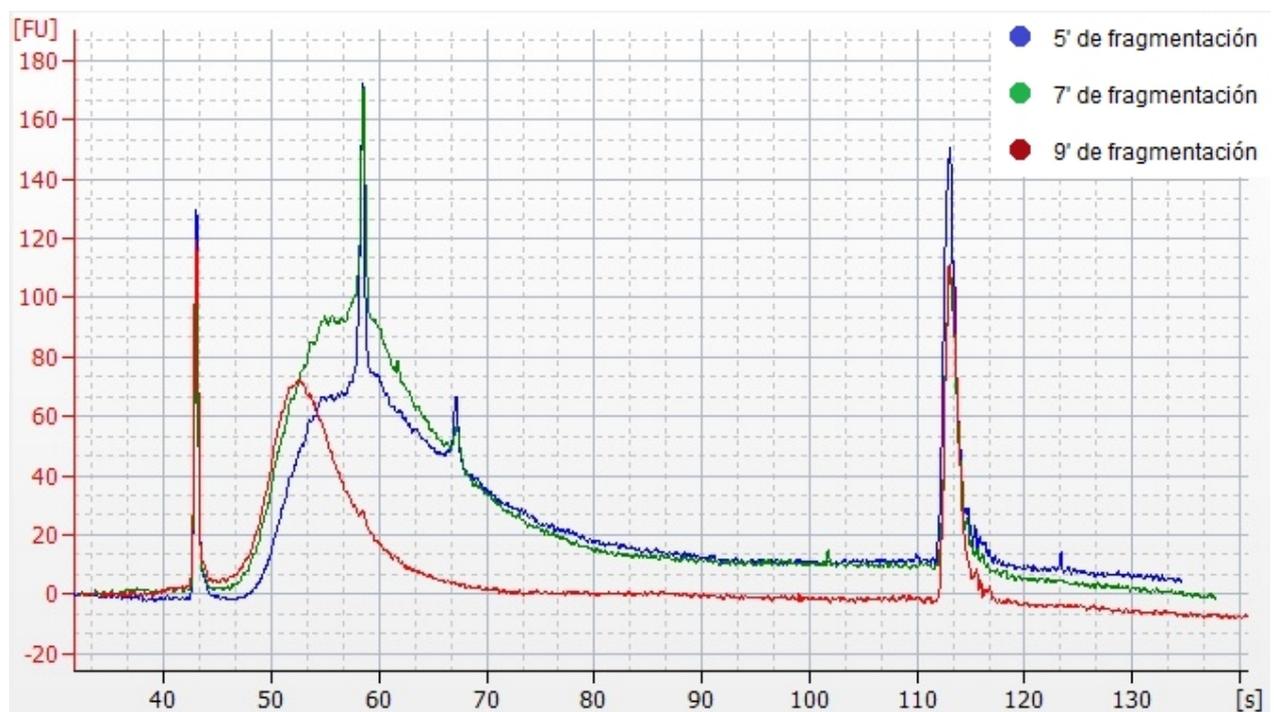


Figura 14 – Fragmentación de los amplicones 4C del potenciador E. La curva azul corresponde a un tiempo de 5', la curva verde a 7' y la curva roja a 9'. El eje de las abscisas se representa el tiempo de migración en la electroforesis capilar, y en el eje de las ordenadas, la señal de fluorescencia.

El perfil de fragmentación a los 5' y 7' fue muy similar, con un rango de tamaño que comenzaba en los 100 pb y contenía un pico en ~ 190 pb. En ambos se observa una espina o “spike” artefacto del análisis. La fragmentación a 9' llevó a la muestra a un rango mucho menor, con un pico en ~ 100 pb.

### Librerías para secuenciar

En base a la prueba de los tiempos de fragmentación, se decidió incubar a 7' los amplicones de cada una de las bibliotecas 4C (del potenciador E y c-Myc). Posteriormente, a los fragmentos se ligaron los adaptadores con un código de barra para poder discriminar entre ambas librerías de secuenciación al final del proceso.

La distribución del tamaño de cada librería de secuenciación, incluyendo los adaptadores que suman 75 pb, se evidenció a través de una electroforesis capilar realizada en Agilent Bioanalyzer 2100. Ambas librerías tenían una distribución homogénea, de entre 150-200 a 500 pb, con un máximo de 220 – 230 pb, y un pico de 45 pb, que posiblemente correspondieran a dímeros de adaptadores (Figura 15).

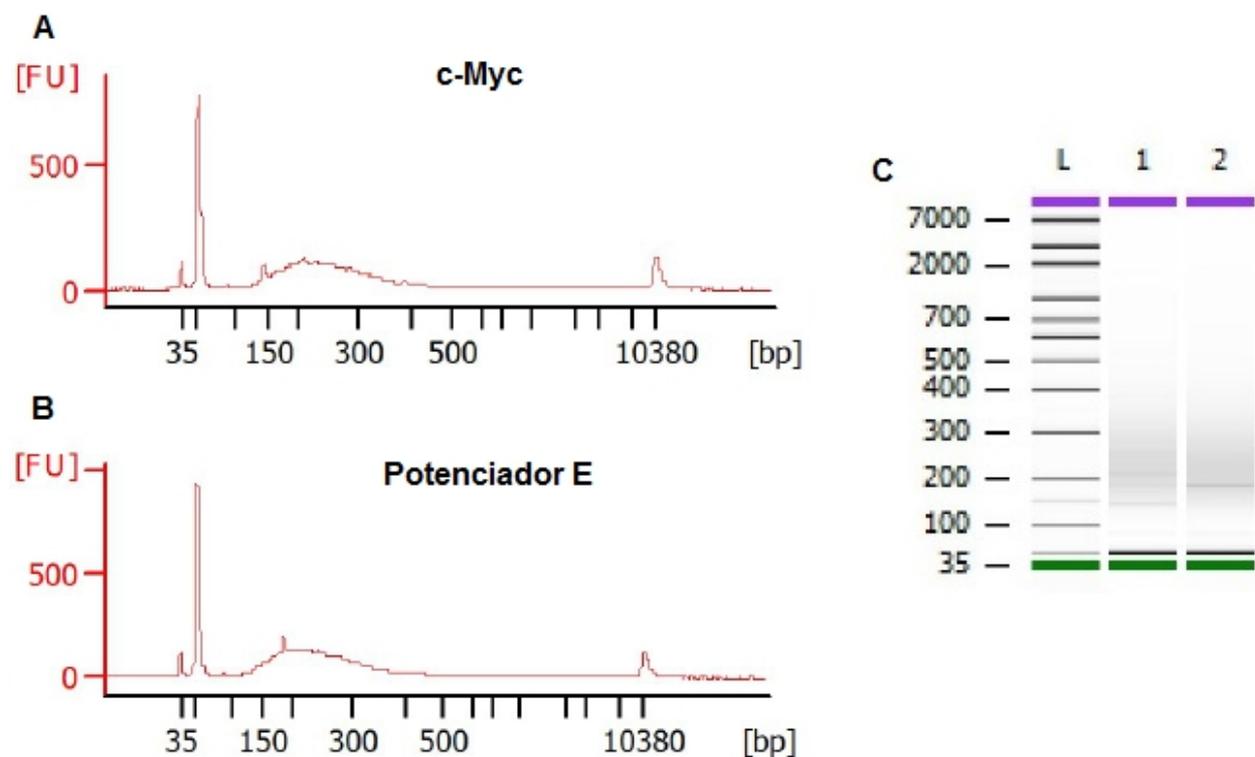
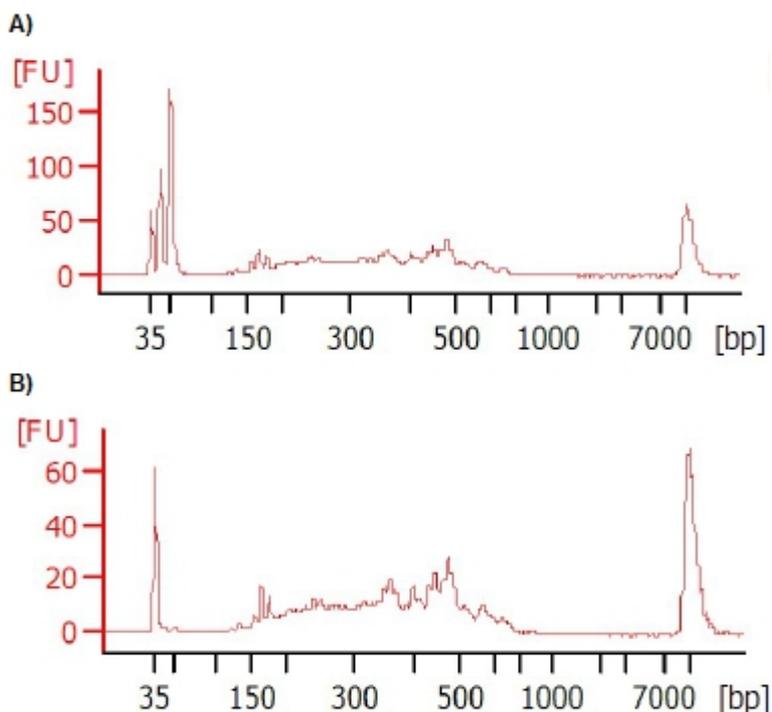


Figura 15 – Perfiles de las librerías de secuenciación de las bibliotecas 4C enriquecidas. A) Electroferograma de la librería del promotor de c-Myc; B) del potenciador E; C) electrofereris virtual de las librerías: carril 1 – ladder, carril 2 – c-Myc, carril 3 – potenciador E. El eje de las abscisas se representa el tamaño en pb, y en el eje de las ordenadas, la señal de fluorescencia.

Para eliminar los probables dímeros de adaptadores, se re-purificaron las librerías con el sistema Agencourt® AMPure® XP con un protocolo modificado donde se agregó 1,6 volúmenes de perlas magnéticas respecto a la muestra. Este procedimiento se utilizó previamente de forma exitosa con una librería de secuenciación de biblioteca 4C ajena a esta tesis (Figura 16).

Figura 16 – Remoción de dímeros de adaptadores de una librería de secuenciación. A) Librería antes de la purificación; B) Misma librería después de la purificación con un volumen 1.6x de perlas magnéticas Agencourt® AMPure® XP. El eje de las abscisas se representa el tamaño en pb, y en el eje de las ordenadas, la señal de fluorescencia.



Las librerías re-purificadas no se sometieron a un nuevo análisis por Bioanalyzer, pero se estimó su concentración a partir del análisis previo a la re-purificación. Se calculó la concentración de cada librería antes de la re-purificación, en un rango comprendido entre 150 y 500 pb, y se estimó un 70% de recuperación después de la re-purificación. Luego se diluyó cada librería para llegar a una concentración 26 pM, y se juntaron en cantidades equimolares para amplificarlas clonalmente en el PCR en emulsión.

### Secuenciación en Ion Torrent

La secuenciación de las bibliotecas 4C del potenciador E y del promotor de c-Myc se llevó a cabo en un Ion 318™ Chip en el equipo Ion PGM™ System (Life Technologies). Se cargó el 46% del chip, lo que equivaldría a 5101643 de ISPs; sin embargo, el número final utilizable fue 3074861 ISPs. Esto se debió a que el 1% de las ISPs estaban vacías, el 16% fue policlonal (durante el PCR en emulsión se combinaron más de una molécula de ADN por ISP), el 2% corresponde a un control interno (Test Fragments) y el 22% se descartaron por baja calidad. El largo promedio de las lecturas obtenidas fue de 146 pb (Figura 17).

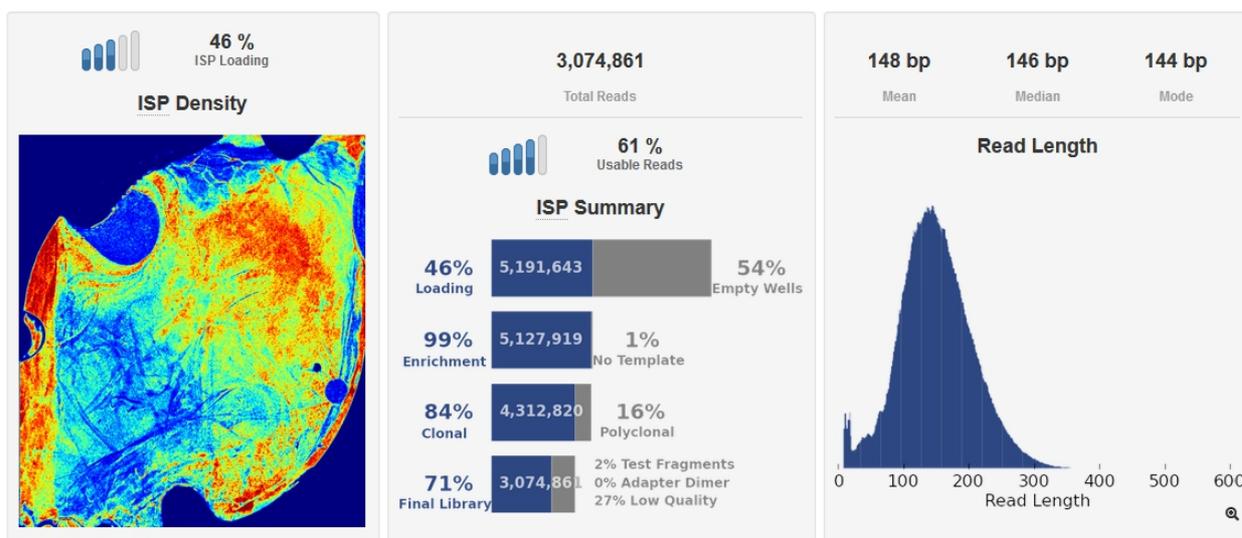


Figura 17 – Primeros resultados de la secuenciación. Izquierda: Densidad de carga de las ISPs en el chip. Medio: Resumen del filtrado de las lecturas. Derecha: Histograma y promedio del largo de las lecturas. Imagen capturada del Torrent Browser asociado al secuenciador.

En la Tabla 5 se pueden apreciar detalles de cada biblioteca 4C secuenciada: el número de bases y de lecturas secuenciadas, así como el rango y el promedio de su largo. El largo de las lecturas fue menor al esperado en la preparación de la librería (fragmentación a 400 pb).

Tabla 5 – Secuenciación de las bibliotecas 4C				
	Bases	Lecturas	Rango largo (pb)	Promedio largo (pb)
Potenciador E	263689900	1713825	8 – 626	153
c-Myc	189476640	1324217	8 – 625	143

### Análisis de datos de secuenciación

Durante el procesamiento de las lecturas se removieron los nucleótidos de baja calidad y se filtraron las lecturas resultantes por el largo (Figura 18). Se alinearon las lecturas procesadas a la secuencia de referencia del genoma humano, obteniéndose en el caso de c-Myc un 97,4% de alineamiento, y en el caso del potenciador E, un 99,7% (Tabla 6).

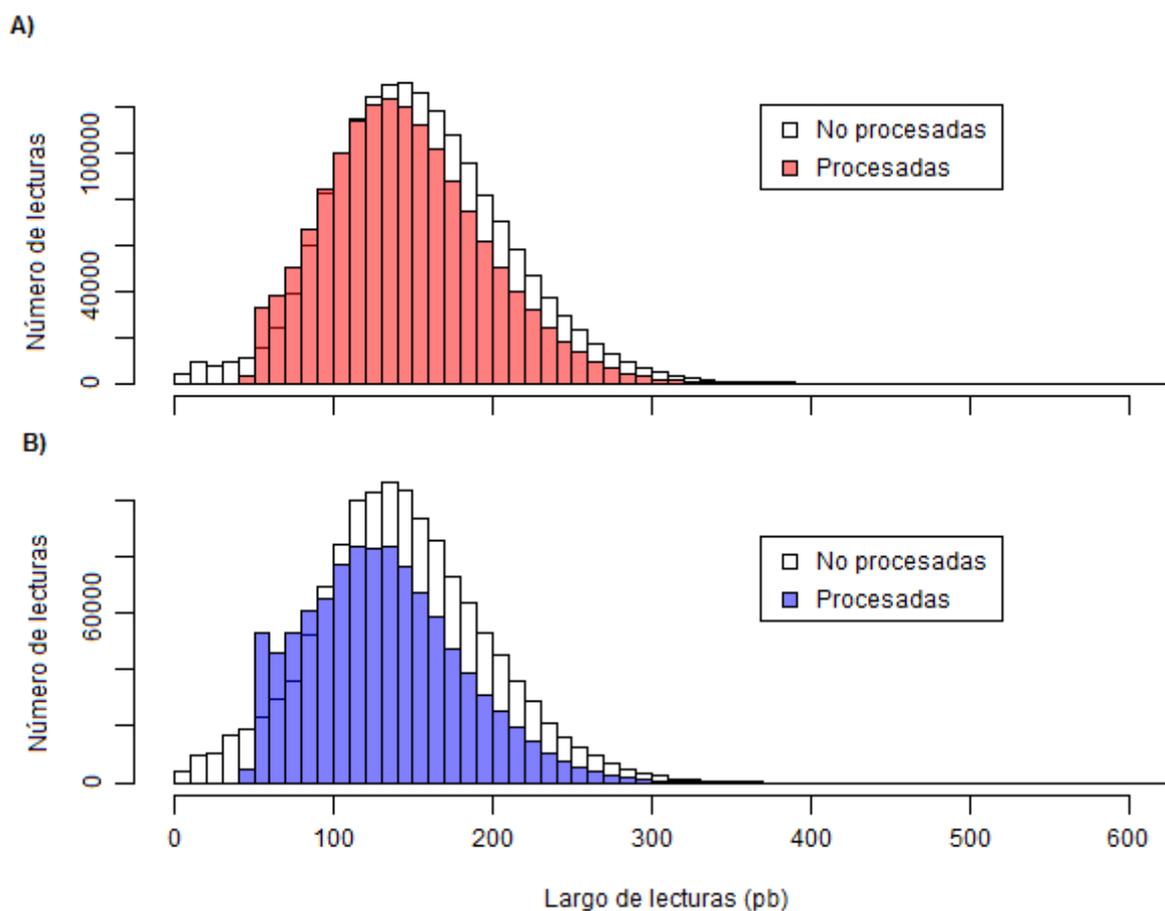


Figura 18 – Distribución del largo de las lecturas de las bibliotecas 4C antes y después de su procesamiento para A) el Potenciador E; B) el promotor de c-Myc.

Tabla 6 – Procesamiento y alineamiento de las lecturas						
	Antes		Después		Alineamiento de las lecturas	
	Lecturas	Largo (pb)	Lecturas	Largo (pb)	Lecturas	Porcentaje (%)
Potenciador E	1713825	8 – 626	1495340	48 - 626	1490738	99,7
c-Myc	1324217	8 – 625	1021207	49 - 625	994507	97,4

Mediante el uso de Samtools y comandos de linux, se manipuló cada archivo 'alineamiento.sam' generado durante el alineamiento para obtener diversos resultados. Se calculó el número de lecturas quiméricas y no quiméricas que alinearon con todo el genoma humano y las que lo hicieron sólo dentro de la región de cada carnada (Tabla 7).

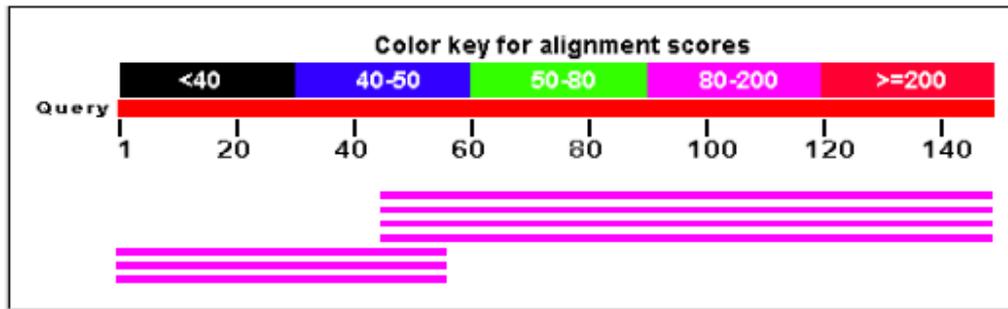
<b>Tabla 7 – Resultados del alineamiento</b>		
	Potenciador E	c-Myc
Lecturas secuenciadas	1495340 (100%)	1021207 (100%)
Lecturas alineadas al genoma	1490738 (99,7%)	994507 (97,4%)
no quiméricas	1184959 (79,2%)	785525 (76,9%)
quiméricas	305779 (20,5%)	208982 (20,5%)
Lecturas no alineadas	4602 (0,3%)	26700 (2,6%)
Lecturas alineadas situadas $\pm$ 2 Kb de la carnada	926240 (61,9%)	764415 (74,8%)
no quiméricas	834814 (55,8%)	681623 (66,7%)
quiméricas	91426 (6,1%)	82792 (8,1%)

Las lecturas quiméricas alineadas contenían fracciones de secuencias que corresponden a diferentes regiones genómicas (Tabla 8). En el caso del potenciador E, el 97.3% de las lecturas quiméricas que alinearon en el genoma contenían dos fracciones, mientras que en el caso de c-Myc el porcentaje fue de 96% (Tabla 8). El porcentaje de lecturas quiméricas conteniendo tres o más fracciones de secuencias es considerablemente menor (Tabla 8).

<b>Tabla 8 – Lecturas quiméricas con distintas fracciones de secuencias (porcentaje respecto al total de quimeras alineadas al genoma)</b>					
Número de fracciones	Dos	Tres	Cuatro	Cinco	Seis
Potenciador E	297653 (97,3%)	7623 (2,5%)	476 (0,2%)	26 (< 0,01%)	1 (< 0,001%)
c-Myc	200682 (96%)	8019 (3,9%)	275 (0,1%)	6 (< 0,01%)	

Para corroborar la existencia de las quimeras, se eligieron dos lecturas para buscar la homología de cada fracción en el genoma humano a través de BLAST. Para los dos casos analizados se comprobó el origen de cada una de las fracciones (Figuras 19 y 20).

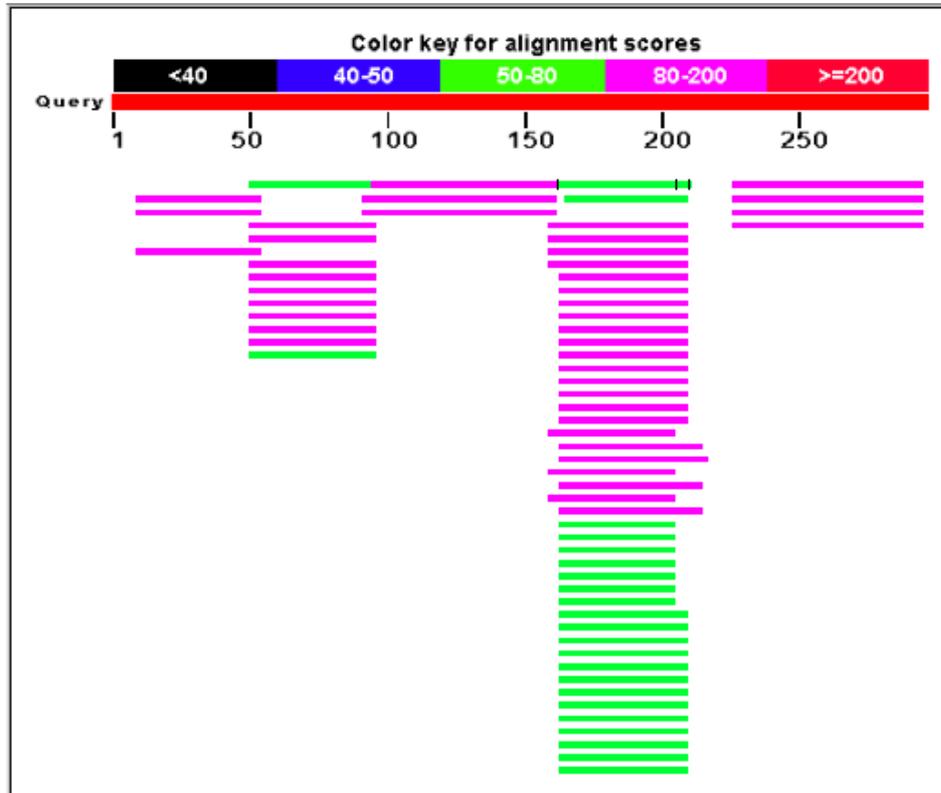
AGCATT TTTT GATGATGAAA GTAGAGGAAA GTATCAAATT CCACCAAAGG ATCAAACCCC  
AACAAATTTA ATTAAGGCTC TTCACGTAGA TGAAATTGAG ATGCTAGACT AGGCAATTTG  
TAATGGTTAG AAAATGCCCT TCTAAGCTT



	Description	Max score	Total score	Query cover	E value	Ident	Accession
<b>Genomic sequences</b> [show first]							
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly CHM1_1.1</a>	193	193	69%	8e-47	100%	<a href="#">NC_018919.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly HuRef</a>	193	193	69%	8e-47	100%	<a href="#">AC_000140.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, GRCh38 Primary Assembly</a>	193	193	69%	8e-47	100%	<a href="#">NC_000008.11</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 11, alternate assembly CHM1_1.1</a>	104	104	37%	4e-20	100%	<a href="#">NC_018922.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 11, alternate assembly HuRef</a>	104	104	37%	4e-20	100%	<a href="#">AC_000143.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 11, GRCh38 Primary Assembly</a>	104	104	37%	4e-20	100%	<a href="#">NC_000011.10</a>

Figura 19 – Homología en el genoma humano de lecturas quiméricas a través de BLAST. Arriba: secuencia de la lectura quimérica de 149 pb con dos fracciones, donde se indica en color rojo el sitio de corte de Mbo I que generaría un fragmento de aproximadamente 50 pb y otro de 100 pb. Abajo: Resultados del BLAST contra el genoma humano, donde el largo de cada una de las barras de colores representa el tamaño de la fracción de la secuencia de interés que alineó contra una secuencia genómica en particular. Se puede apreciar que los tamaños de las barras son similares a los esperados por la digestión de la lectura. También se muestran los sitios de homología en el genoma (matches), que en este caso corresponden a los cromosomas 8 y 11.

CAGTGGCATG ATCTCTGCTG GGCCATGTTA CCTGTGGCCT TGGGATATTT GGATCACTTG  
 GGCCCTGGAG GTCAAGGCTG CAGTGAGCTAT GATCAGTGAC GGTGGGAAC TAACTCTGGG  
 AACTATAGC TTCTTTAAAA GTGGAAGGAT CTAGTGATCC GACTCTGTCA CCCAGGCTGG  
 AGTGCAGTGG CACGATCTTA GCTTGCTGGG CTCTGTTGGG ATGGGATCTG TGATGCTTCT  
 CCTATAAACT TGTTCACAG AAAACCAGCT TACCCAGGTA TAGGCTCAGT TTTA



Description	Max score	Total score	Query cover	E value	Ident	Accession
<b>Genomic sequences [show first]</b>						
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 8, alternate assembly CHM1_1.1</a>	130	1.810e+05	56%	1e-27	100%	<a href="#">NC_018919.2</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 8, alternate assembly HuRef</a>	130	1.783e+05	56%	1e-27	100%	<a href="#">AC_000140.1</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 8, GRCh38 Primary Assembly</a>	130	1.817e+05	56%	1e-27	100%	<a href="#">NC_000008.11</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome X, alternate assembly CHM1_1.1</a>	128	2.016e+05	60%	5e-27	100%	<a href="#">NC_018934.2</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome X, alternate assembly HuRef</a>	128	1.660e+05	60%	5e-27	100%	<a href="#">AC_000155.1</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome X, GRCh38 Primary Assembly</a>	128	2.030e+05	60%	5e-27	100%	<a href="#">NC_000023.11</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome Y, GRCh38 Primary Assembly</a>	128	27348	59%	5e-27	100%	<a href="#">NC_000024.10</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 13, alternate assembly CHM1_1.1</a>	86.1	1.028e+05	52%	3e-14	100%	<a href="#">NC_018924.2</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 18, alternate assembly CHM1_1.1</a>	86.1	96492	37%	3e-14	100%	<a href="#">NC_018929.2</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 18, alternate assembly HuRef</a>	86.1	95076	37%	3e-14	100%	<a href="#">AC_000150.1</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 13, alternate assembly HuRef</a>	86.1	1.023e+05	50%	3e-14	100%	<a href="#">AC_000145.1</a>
<input type="checkbox"/> <a href="#">Homo sapiens chromosome 13, GRCh38 Primary Assembly</a>	86.1	1.045e+05	52%	3e-14	100%	<a href="#">NC_000013.11</a>

Figura 20 – Homología en el genoma humano de lecturas quiméricas a través de BLAST. Arriba: secuencia de la lectura con seis fracciones, donde se indica en color rojo los sitios de corte de Mbo I que generaría fragmentos de alrededor de 30, 40, 60, y 70 pb. Abajo: Resultados del BLAST contra el genoma humano, donde cada una de las barras de colores representa la fracción de la secuencia de interés que alineó contra una secuencia genómica en particular. Se puede apreciar que los tamaños de las barras son similares a los esperados por la digestión de la lectura. También se muestran los primeros matches del alineamiento, que incluyen al cromosoma 8, 13, 18, X e Y.

## Búsqueda de picos 4C

Para identificar picos 4C o regiones genómicas enriquecidas en lecturas, se utilizó el programa MACS. Se encontraron 3239 picos para el caso de c-Myc, 3625 picos para el caso del Potenciador E, y 88 picos en común a ambas carnadas. Los picos del potenciador E tenían un rango de extensión de 137 a 3152 pb, mientras que los de c-Myc el rango era de 202 a 3857 pb.

Se estudió las enfermedades y desórdenes en los que estaban involucrados los genes más cercanos a los 88 picos 4C en común, así como su función molecular y celular, a través de Ingenuity Pathway Analysis de QIAGEN (IPA®, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)). Se encontraron 111 genes cercanos, de los cuales 56 están involucrados en cáncer (Tabla 9) y varios tienen funciones relacionadas con muerte celular y sobrevivencia, desarrollo, crecimiento y proliferación celular (Tabla 9).

Tabla 9 – Ontología de los genes más cercanos a los picos 4C en común a ambas carnadas		
<b>Diseases and Disorders</b>		
Name	p-value	# Molecules
Immunological Disease	3,42E-04 - 4,67E-02	6
Neurological Disease	3,42E-04 - 3,78E-02	9
Skeletal and Muscular Disorders	3,42E-04 - 1,58E-02	6
Cancer	5,18E-03 - 4,67E-02	56
Cardiovascular Disease	5,29E-03 - 4,67E-02	10
<b>Molecular and Cellular Functions</b>		
Name	p-value	# Molecules
Cellular Assembly and Organization	2,77E-05 - 4,67E-02	8
Cellular Function and Maintenance	7,60E-04 - 4,65E-02	10
Cell Death and Survival	2,08E-03 - 4,91E-02	12
Cellular Development	2,78E-03 - 4,67E-02	10
Cellular Growth and Proliferation	2,78E-03 - 4,27E-02	8

En la Tabla 9 se muestra el análisis de los genes más cercanos a los 88 picos 4C en común entre el potenciador E y el promotor de c-Myc. Se puede apreciar la lista de enfermedades y desórdenes a los que están asociados los genes, su valor de significancia y el número de genes involucrados, así como también sus funciones celulares y moleculares, su valor de significancia y el número de genes involucrados.

También se obtuvieron los 10 picos 4C con mayor número de lecturas dentro de los datos 4C para cada carnada (Tabla 10). No se incluyeron los picos localizados en las carnadas; aunque se observó la presencia picos 4C en el promotor de c-Myc, en datos del potenciador E, y picos en cercanía del potenciador E, en datos de c-Myc.

<b>Tabla 10 – Picos 4C con mayor número de lecturas para cada carnada</b>				
<b>Potenciador E como carnada</b>				
<b>Localización genómica (coordenadas)</b>	<b>Banda citogenética</b>	<b>Gen/genes más cercano/s</b>	<b>Número de secuencias</b>	<b>Extensión del pico (pb)</b>
chr2:29105926-29106817	2p23.2	5' de WDR43 y TRMT61B	867	892
chr19:50584756-50585523	19q13.33	5' de genes SNAR	815	768
chr10:42384406-42385812	10q11.21 (centrómero)	3 de LOC441666	776	1407
chr10:83506812-83507591	10q23.1	5' de NRG3	718	780
chr13:41836876-41837699	13q14.11	MTRF1	691	824
chr20:57161721-57162472	20q13.32	APCDD1L	685	752
chr1:121483960-121485528	1p11.2 (centrómero)	EMBP1	667	1569
chr6:166304424-166305032	6q27	3' de LINC00473	658	609
chr10:43596653-42598486	10q11.21 (centrómero)	3' de CCNYL2	633	1834
chr8:105625797-105626614	8q22.3	5' de LRP12	631	818
<b>Promotor de c-Myc como carnada</b>				
<b>Localización genómica (coordenadas)</b>	<b>Banda citogenética</b>	<b>Gen/genes más cercano/s</b>	<b>Número de secuencias</b>	<b>Extensión del pico (pb)</b>
chr18:107870-110634	18p11.32	ROCK1P1	690	690
chr17:54986350-54986985	17q22	TRIM25	393	791
chr13:26223318-26224108	13q12.13	ATP8A2	383	791
chr16:87897815-87898481	16q24.2	SLC7A5	348	667
chr4:169998150-169998783	4q32.3	5' de CBR4, 3' de SH3RF1	344	634
chr11:114102712-114103383	11q23.2	ZBTB16	339	672
chr2:140237744-140238410	2q22.1	3' de YY1P2	329	667
chr20:46586753-46587348	20q13.13	LINC01522	325	596
chr17:33861248-33861969	17q12	3' de SLFN14	324	722
chr1:36980852-36981407	1p34.3	5' de CSF3R	323	556

Para los genes cercanos a los picos 4C más abundantes para el potenciador E y para el promotor de c-Myc, se aplicó el mismo análisis que para las intersecciones, a través de Ingenuity Pathway Analysis. De 22 genes cercanos, varios tienen funciones relacionadas con desarrollo y movimiento celular (Tabla 11).

**Tabla 11 – Ontología de los genes más cercanos a los picos 4C más abundantes**

<b>Diseases and Disorders</b>		
Name	p-value	# Molecules
Cancer	1,49E-03 - 4,73E-02	20
Cardiovascular Disease	1,49E-03 - 4,80E-02	5
Connective Tissue Disorders	1,49E-03 - 2,65E-02	3
Developmental Disorder	1,49E-03 - 2,21E-02	6
Hematological Disease	1,49E-03 - 4,37E-02	6
<b>Molecular and Cellular Functions</b>		
Name	p-value	# Molecules
Cell Morphology	2,13E-05 - 3,51E-02	3
Amino Acid Metabolism	1,49E-03 - 8,90E-03	2
Cell Cycle	1,49E-03 - 4,09E-02	2
Cell Death and Survival	1,49E-03 - 3,80E-02	6
Cellular Development	1,49E-03 - 4,52E-02	4

En la Tabla 11 se muestra el análisis de los genes más cercanos a los picos 4C más abundantes para el potenciador E y para el promotor de c-Myc. En la parte superior se detalla la lista de enfermedades y desórdenes a los que están asociados los genes, su valor de significancia y el número de genes involucrados, mientras que en la parte superior se detallan las funciones celulares y moleculares de los genes, su valor de significancia y el número de genes involucrados.

## **Análisis de picos 4C**

### Distribución y posicionamiento de picos 4C en el genoma

Para estudiar la distribución y posicionamiento de todos los picos 4C en el genoma y estructuras genómicas, se empleó el programa CEAS, quien compara las regiones alineadas con regiones mapeables del genoma (“background”). Respecto a la distribución genómica, el potenciador E presentó un mayor porcentaje de picos 4C en los cromosomas 1, 5, 8, 9, 10 y 21, mientras c-Myc lo tiene en los cromosomas 1, 7, 8, 9, 15, 16, 17, 19 y 20 (Figura 21).

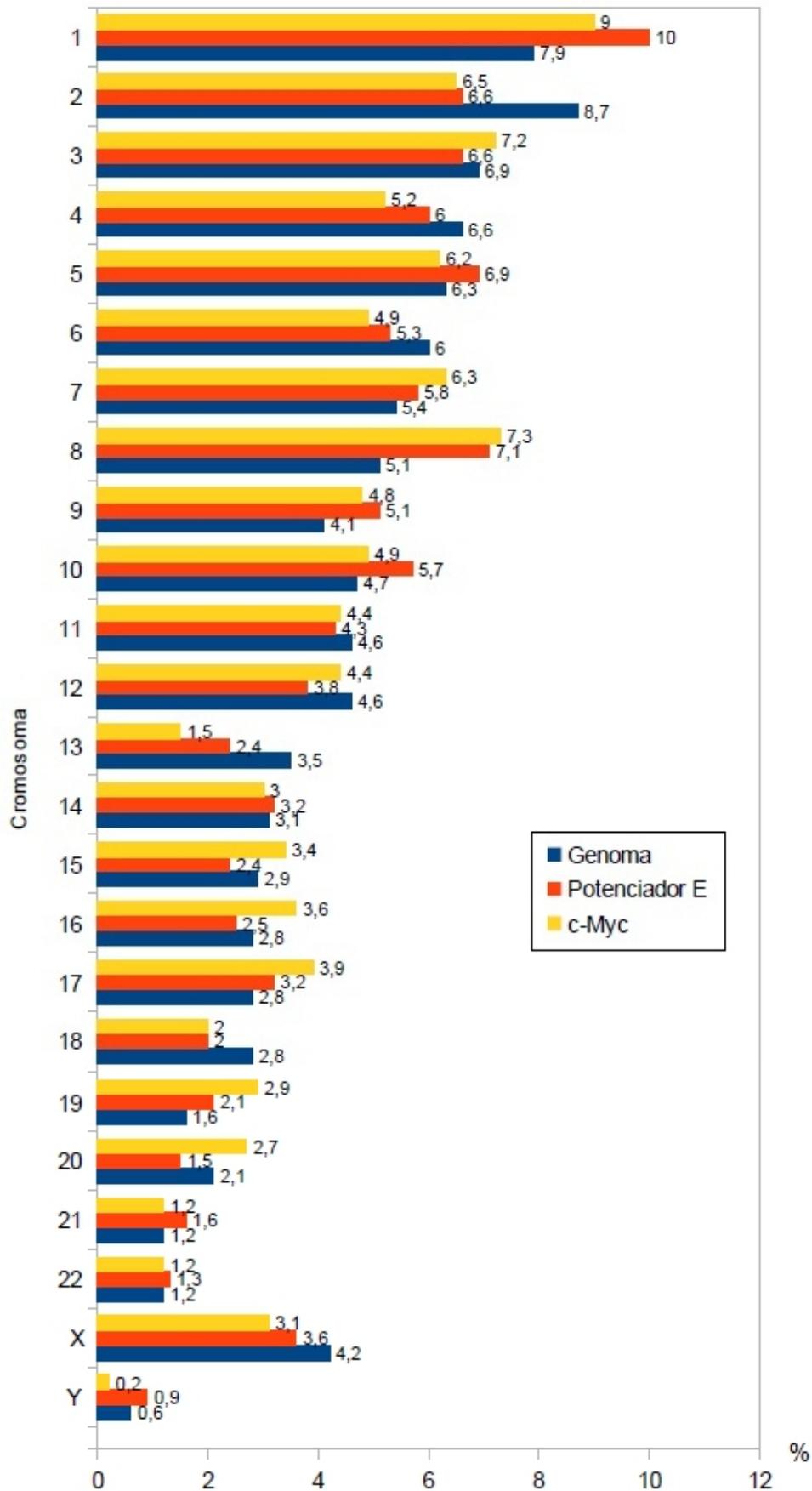


Figura 21 – Distribución de picos 4C en el genoma. El largo de las barras representa el porcentaje de picos 4C en cada uno de los cromosomas; barra amarilla representa picos 4C de c-Myc, barra roja del potenciador E, y barra azul las regiones mapeables del genoma (“background”).

En el cromosoma 8, donde se encuentran las carnadas, se presenta la mayor diferencia de porcentaje en relación al background: 7,1% en el potenciador y 7,3% en c-Myc vs 5,1%.

En la Figura 22 se muestra la distribución de los picos 4C en características genómicas tales como promotores y regiones río abajo de TTS (con distancias relativas), UTRs, exones, intrones y regiones intergénicas distales para cada experimento 4C. Las regiones intergénicas distales corresponden a regiones del genoma que no pertenecen a ninguna de las otras categorías.

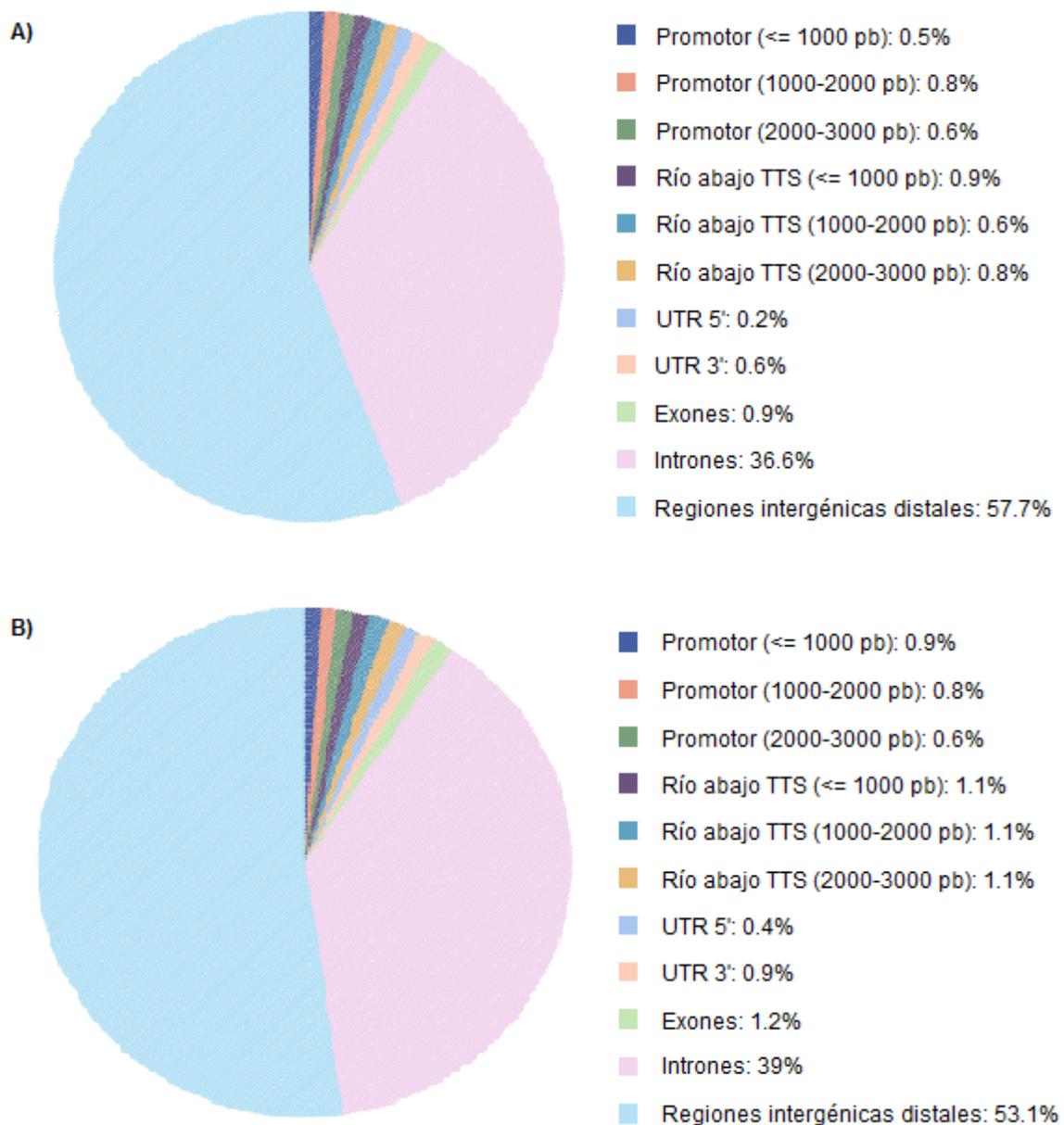


Figura 22 – Distribución de los picos 4C en características genómicas. Izquierda: Representación gráfica de la distribución; Derecha: Porcentaje de picos 4C pertenecientes a cada categoría. A) picos 4C del potenciador E; B) picos 4C del promotor de c-Myc.

También se estudió el promedio de picos 4C en un metagen, una estructura génica de 3kb definida arbitrariamente para comparar con el background de aquellos genes del genoma con el mismo largo. El potenciador E y c-Myc presentan un enriquecimiento en las regiones 5' y 3' respecto al metagen, en comparación con el background (Figura 23).

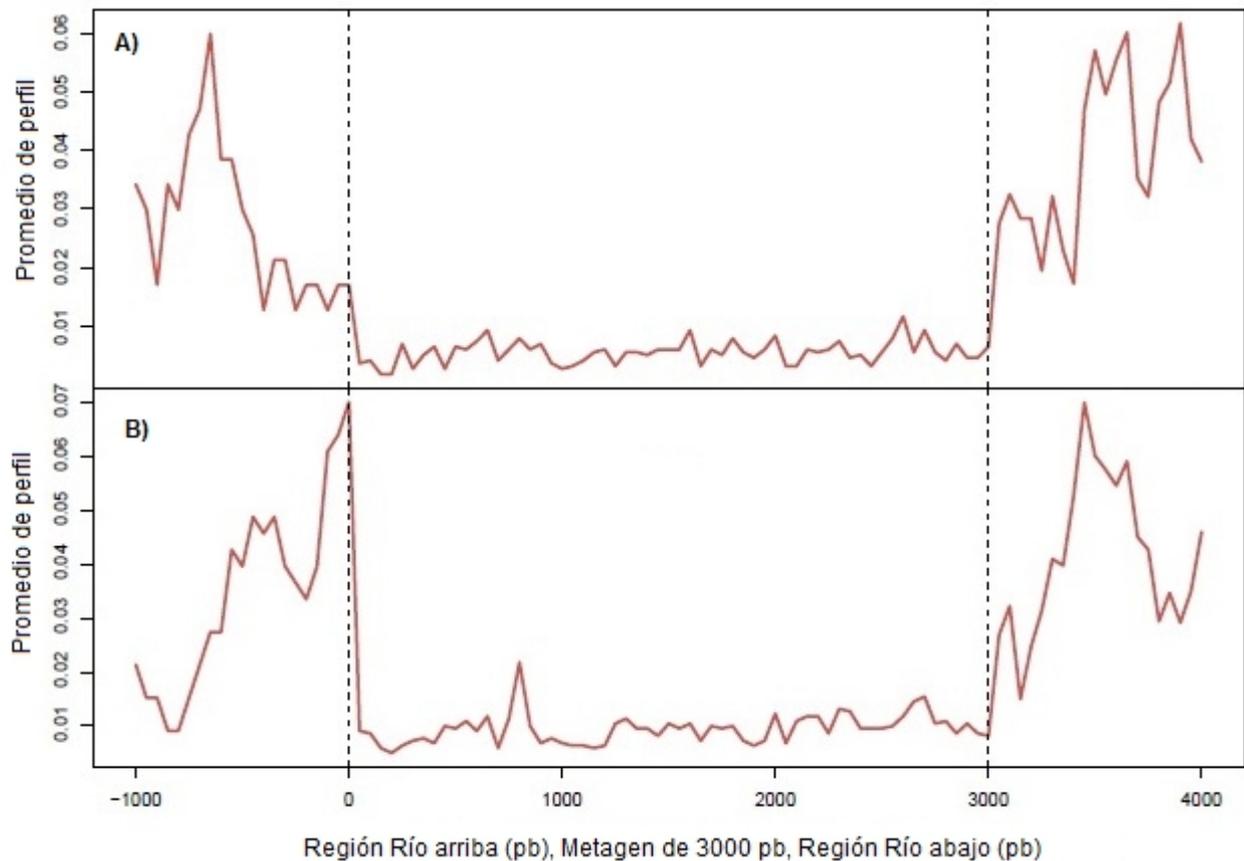


Figura 23 – Promedio de picos 4C respecto a un metagen de 3000 pb y en regiones río arriba y abajo del mismo. A) Potenciador E; B) c-Myc.

Para estudiar el posicionamiento de los picos 4C en las características genómicas antes mencionadas, también se utilizó a CEAS. En las regiones promotoras, c-Myc mostró un porcentaje de picos 4C similar al esperado para el background, mientras el potenciador E mostró un porcentaje menor (Figura 24A).

En las regiones río abajo de genes, el potenciador E presentó un porcentaje de picos 4C similar al esperado para el background, mientras que para c-Myc el porcentaje fue mayor, incrementándose a medida que la región se alejaba del sitio de terminación de la transcripción o TTS (Transcription Termination Site) de genes (Figura 24A). No hubo diferencias en la distribución de picos 4C en las estructuras génicas respecto al background (Figura 24B).

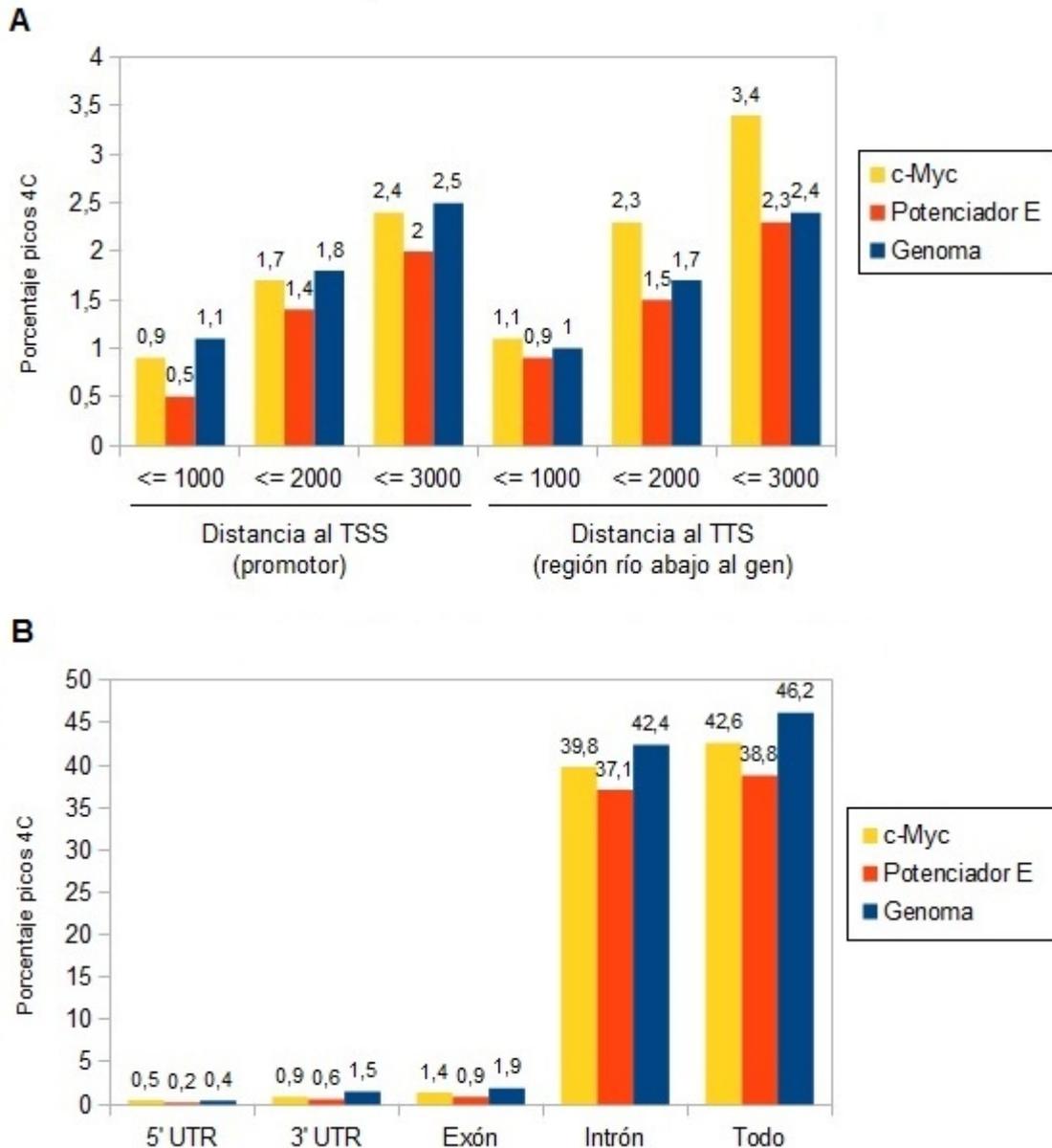


Figura 24 – Distribución de picos 4C en estructuras génicas. A) En promotores y regiones río abajo de genes; B) En cuerpos génicos. La barra amarilla representa el porcentaje de picos 4C de c-Myc, la barra roja el del potenciador E, y la barra azul las regiones mapeables del genoma (“background”).

### Búsqueda de marcadores funcionales en coincidencia con las zonas de interacción obtenidas por 4C

Para conocer la función o el vínculo que tienen los datos 4C que interactúan con el potenciador E o el promotor de c-Myc, se cruzaron los picos 4C con datos de ChIP-Seq y DNAsa-Seq que indicaran la presencia de elementos genómicos activos y participes de interacciones de largo alcance.

Los genes regulados por el potenciador E se identificaron al intersectar picos 4C con datos de CHIP-Seq para H3K4me3, un marcador de regiones promotoras (Figura 25). Se encontraron 178 genes asociados a 156 intersecciones, algunas de las cuales estaban asociadas o rodeadas, río arriba y río abajo, por más de un gen.

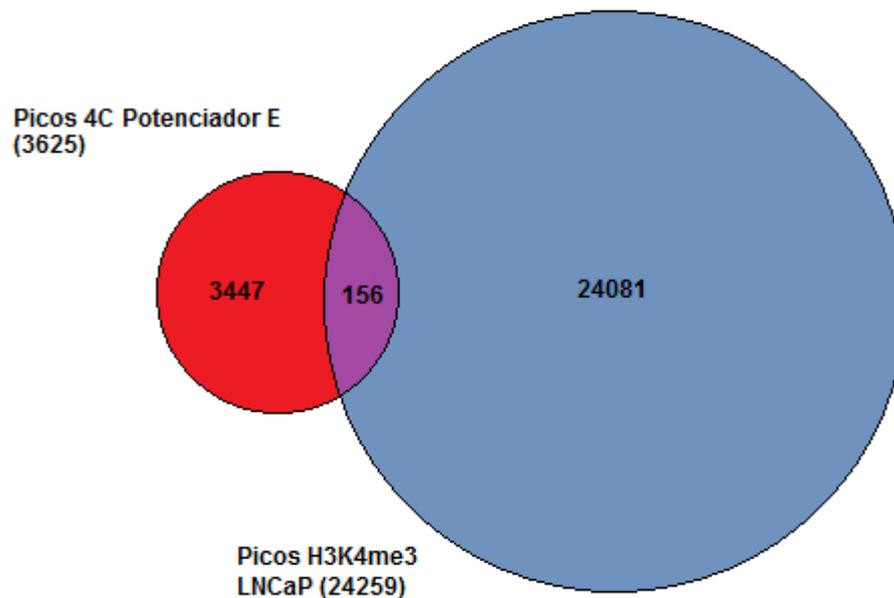


Figura 25 – Intersección de picos 4C del potenciador E con marcadores de promotores. El diagrama de Venn representa la intersección de picos 4C del potenciador E con picos de CHIP de H3K4me3, marcador de regiones promotoras en LNCaP.

Solo 22 genes tenían una intersección en la región comprendida entre -2000 pb y + 500 pb respecto a su TSS, para los cuales también se buscó colocación de las intersecciones con datos de CHIP-Seq para algunas proteínas de interés (cromatina abierta, factores de transcripción, formación de lazos de cromatina), DNase-Seq (cromatina abierta) y su expresión en Oncomine (Tabla 12).

Tabla 12 – Genes con picos 4C del potenciador E y H3K4me3 en su región promotora									
Gen	Localización	TSS	CTCF	TCF4	STAG1	RAD21	PoI II	DHS	Oncomine
AAMP	2q35	440							
ACBD3	1q42.12	-719							
ADAM23	2q33.3	-531							
ADAMTS6	5q12.3	-612							
BCL9	1q21.2	-279							
CCNY	10p11.21	-1141							
COG7	16p12.2	-42							
DSE	6q22.1	-1141							
HPSE	4q21.23	-642							
IFITM3	11p15.5	-206							
MTRF1	13q14.11	446							
c-MYC	8q24.21	-1305							
NSUN5	7q11.23	-821							
PEX1*	7q21.2	-559							
PNKD	2q35	-662							
RBM48*	7q21.2	317							
RNGTT	6q15	-171							
RPL41	12q13.2	-1382							
RTF1	15q15.1	-1034							
RYBP	3p13	-1222							
SMARCC1	3p21.31	-1424							
ZNF643	1p34.2	116							

\* Los genes PEX1 y RBM48 comparten la misma intersección picos 4C x H3K4me3

En la Tabla 12 se detalla la información referente a los 22 genes resultantes de cruzar datos 4C con ChIP y DNAsa-Seq: nombre del gen, localización cromosómica, y distancia de la intersección 4C y H3K4me3 respecto a su TSS. También se indica la ausencia (naranja) o presencia (verde) de colocación con datos ChIP/DNAsa-Seq de interés así como sus niveles de sobre (rojo) o subexpresión (azul). En color blanco se presentan los genes sin expresión diferencial, y en gris los casos en que no hubo medición de la expresión en ese estudio).

Un resumen de las colocalizaciones entre datos ChIP – ChIP/DNAsa puede apreciarse en la Tabla 13, donde se indica el porcentaje de genes que coincidían con señales particulares de la cromatina, y el porcentaje de genes con colocalizaban varias señales.

<b>Tabla 13 – Marcadores funcionales en picos 4C del potenciador E (H3K4me3)</b>						
H3K4me3 x Potenciador E	CTCF(%)	TCF4 (%)*	STAG1 (%)	RAD21 (%)	Pol II (%)	DHS (%)
22	7 (32)	3 (14)	17 (77)	14 (63)	17 (77)	14 (63)
CTCF		2 (9)	6 (27)	6 (27)	6 (27)	5 (23)
TCF4*			3 (14)	3 (14)	3 (14)	3 (14)
STAG1				14 (63)	16 (73)	11 (50)
RAD21					12 (54)	10 (45)
Pol II						11 (50)
DHS						

A partir de estos 22 genes, se seleccionaron como candidatos a aquellos cuyas intersecciones asociadas colocalizaran con los datos ChIP y DNAsa-Seq utilizados, y que estuviesen sobreexpresados en Oncomine o que la literatura indicara una posible vinculación con cáncer (Tabla 14).

<b>Tabla 14 – Genes seleccionadas como candidatos regulados por el potenciador E</b>
<p><u>ACBD3 (acyl-CoA binding domain containing 3)</u>            Proteína involucrada en la estructura y función del aparato de Golgi, así como también en esteroidogenesis dependiente de cAMP, apoptosis y homeostasis del hierro (revisado en Fan et al. 2010). Análisis de microarreglos publicados en trabajos sobre perfiles de expresión en distintos tipos de cáncer, mostraron que ACBD3 está involucrada en el control del ciclo celular y que es regulada por la proteína de retinoblastoma (pRB), un supresor de tumores (Finocchiaro et al. 2005).</p>
<p><u>BCL9 (B-cell CLL/lymphoma 9)</u>            El gen BCL9 fue descubierto al clonar una translocación de un paciente con leucemia linfoblástica aguda (Willis et al. 1998). Promueve la proliferación, migración, invasión y potencial metastásico de células de mieloma múltiple y CRC (Mani et al. 2009). La sobreexpresión de BCL9 coincide con la activación aberrante de la vía Wnt en CRC y cáncer de mama (Adachi et al. 2004; Sakamoto et al. 2007), y es requerido para la transcripción de genes blancos de la vía Wnt mediada por <math>\beta</math>-catenina en CRC (de la Roche et al. 2008; Takada et al. 2012).</p>
<p><u>CCNY (cyclin Y)</u>            Ciclina que controla el ciclo celular y regula a quinasas dependientes de ciclina. Sobreexpresado en non-small-cell lung cancer (Yue et al. 2011). CCNY forma un complejo con la quinasa PFTK1, frecuentemente sobreexpresada en carcinoma hepatocelular, que activa a Rho GTPasas para promover la vía Wnt no canónica (Sun et al. 2014).</p>
<p><u>HPSE (Heparanase)</u>            Proteína que cliva el heparansulfato de membrana basal y matriz extracelular para permitir el movimiento celular. Sobreexpresado en cáncer de mama (Tang et al. 2013), ovarios (Zhang et al. 2013), pulmón (Song et al. 2008) y próstata (Lerner et al. 2008). Su sobreexpresión en cáncer de células no pequeñas de pulmón sugiere una correlación con metástasis (Zhou et al., 2008).</p>

**Tabla 14 – Genes seleccionadas como candidatos regulados por el potenciador E (continuación)**

RTF1 (Rtf1 Paf1/RNA polymerase II complex component homolog (*S. cerevisiae*))

Subunidad de Paf1C (homólogo a Paf1 en levadura), un complejo que colocaliza con RNA Pol II durante la elongación de la transcripción y que es requerido para la modificación de histonas. Se ha propuesto que RTF1 sería el componente primario que coordina dichas modificaciones, influyendo en la transcripción de genes en levadura (Warner et al. 2007). Aunque sería subexpresado en Oncomine (Tabla 12), se seleccionó porque podría estar involucrado en la transcripción de genes supresores de tumores.

SMARCC1 (SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily c member 1)

Miembro de la familia de proteínas SWI/SNF, con actividad helicasa y ATPasa que regularían la transcripción por alteración de estructura de la cromatina. Es sobreexpresado en PCa (Heebøll et al. 2008), CRC (Andersen et al. 2009) y neoplasia cervical intraepitelial (Shadeo et al. 2008). Se observó una correlación positiva de su expresión con desdiferenciación, progresión, metástasis y tiempo de recurrencia en PrC (Heebøll et al., 2008), y que su metilación controlaría la expresión de genes de la vía c-Myc (Wang et al. 2014b). Por otro lado, se propuso a SMARCC1 como un supresor de tumores, pues su re-expresión en células derivadas de cáncer defectivas para su gen (carcinoma de ovario y colon), llevó a una reducción de colonias (DeIBove et al. 2011).

ZNF643 (ZFP69 zinc finger protein B)

Regulador de la transcripción. No se encontró literatura que lo relacionara con cáncer, pero fue seleccionado por su sobreexpresión en muestras analizadas en Oncomine (Tabla 12).

Para encontrar posibles elementos reguladores de la expresión de c-Myc, se interseccionaron picos 4C con datos de ChIP-Seq para H3K4me1, H3K27ac y p300, marcadores de elementos potenciadores. No se encontró intersección entre los picos 4C de c-Myc y los tres marcadores; pero sí se encontraron 86 intersecciones de picos 4C con H3K4me1 y H3K27ac (Figura 26; Tabla 15), y 8 intersecciones de picos 4C y p300 (Tabla 16).

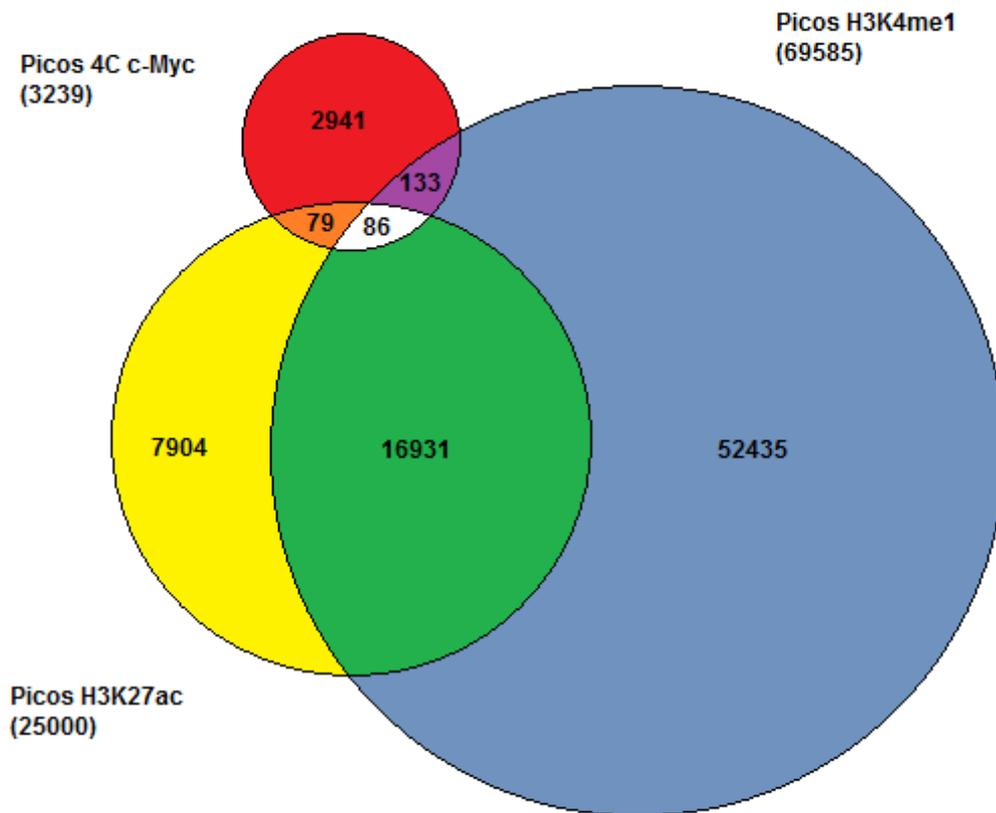


Figura 26 – Intersección de picos 4C de c-Myc con marcadores de potenciadores. El diagrama de Venn representa las intersecciones del picos 4C del promotor de c-Myc con picos de CHIP de H3K4me1 y H3K27ac, marcadores de potenciadores en LNCaP.

Resúmenes de las colocalizaciones entre datos CHIP – CHIP-DNAsa se aprecian en las Tablas 15 y 16 , donde se indica el porcentaje de genes que coincidían con señales particulares de la cromatina, y el porcentaje de genes con colocalizaban varias señales.

Tabla 15 – Marcadores funcionales en picos 4C de c-Myc (H3K4me1 y H3K27ac)						
H3K4me1 x H3K27ac x Myc	CTCF (%)	TCF4 (%)	STAG1 (%)	RAD21 (%)	Pol II (%)	DHS (%)
86	10 (12)	5 (6)	21 (24)	19 (22)	28 (32)	54 (63)
CTCF		0 (0)	6 (7)	5 (6)	4 (5)	5 (6)
TCF4			2 (2)	2 (2)	2 (2)	3 (3)
STAG1				17 (20)	9 (10)	17 (20)
RAD21					7 (8)	12 (14)
Pol II						12 (14)
DHS						

Tabla 16 – Marcadores funcionales en picos 4C de c-Myc (p300)								
Myc x p300	H3K4me1 (%)	H3K27ac (%)	CTCF (%)	TCF4 (%)	STAG1 (%)	RAD21 (%)	Pol II (%)	DHS (%)
8	6 (75)	0 (0)	2 (25)	1 (13)	2 (25)	2 (25)	7 (88)	5 (63)
H3K4me1		-	1 (13)	1 (13)	1 (13)	1 (13)	6 (75)	3 (38)
CTCF				0	2 (25)	2 (25)	2 (25)	2 (25)
TCF4					0	0	1 (13)	1 (13)
STAG1						2 (25)	2 (25)	2 (25)
RAD21							2 (25)	2 (25)
Pol II								4 (50)
DHS								

A partir de estos resultados, se seleccionaron como candidatos a elementos reguladores a las intersecciones que colocalizaran con datos de ChIP y/o DNAsa – Seq para algunas proteínas de interés (cromatina abierta, factores de transcripción, formación de lazos de cromatina) así como DNAsa-Seq (cromatina abierta) (Tabla 17).

Tabla 17 – Regiones seleccionadas como potenciadores candidatos a interactuar con c-Myc									
	H3K4me1	H3K27ac	CTCF	TCF4	STAG1	RAD21	Pol II	DHS	p300
1p34.3	Verde	Naranja	Verde	Naranja	Verde	Verde	Verde	Verde	Verde
1p11.2	Verde	Naranja	Naranja	Naranja	Naranja	Naranja	Verde	Naranja	Verde
1q25.3	Verde	Verde	Verde	Naranja	Verde	Verde	Verde	Verde	Naranja
3p25.3	Verde	Verde	Verde	Naranja	Verde	Verde	Verde	Verde	Naranja
9p23	Verde	Verde	Naranja	Verde	Verde	Verde	Verde	Naranja	Naranja
15q15.3	Verde	Verde	Naranja	Naranja	Verde	Verde	Verde	Verde	Naranja
19p13.3	Verde	Verde	Naranja	Naranja	Verde	Verde	Verde	Naranja	Naranja
21q22.3	Verde	Verde	Naranja	Naranja	Verde	Verde	Verde	Verde	Naranja

En la Tabla 17 se indica como referencia localización cromosómica de los potenciadores seleccionados, y la colocalización (verde) o no (naranja) con datos ChIP/DNAsa-Seq de interés.

## Visualización de los picos 4C

### Con IGV

En las Figuras 27 y 28 se muestra de forma panorámica la distribución de los picos 4C en el genoma, tanto del potenciador E como el promotor de c-Myc, y de los datos de ChIP y DNAsa-Seq utilizados en esta tesis. También se explora en mayor profundidad algunas de las interacciones candidatas identificadas, mostrando la colocalización de datos utilizados.



Figura 27 – Colocalización de un pico 4C de c-Myc y marcadores funcionales. Arriba: Panorámica de la distribución de los picos 4C en todo el genoma, acompañado de datos ChIP y DNAsa-Seq. Abajo: Magnificación de la interacción 15q15.3 – c-Myc y solapamiento con datos ChIP y DNAsa-Seq.

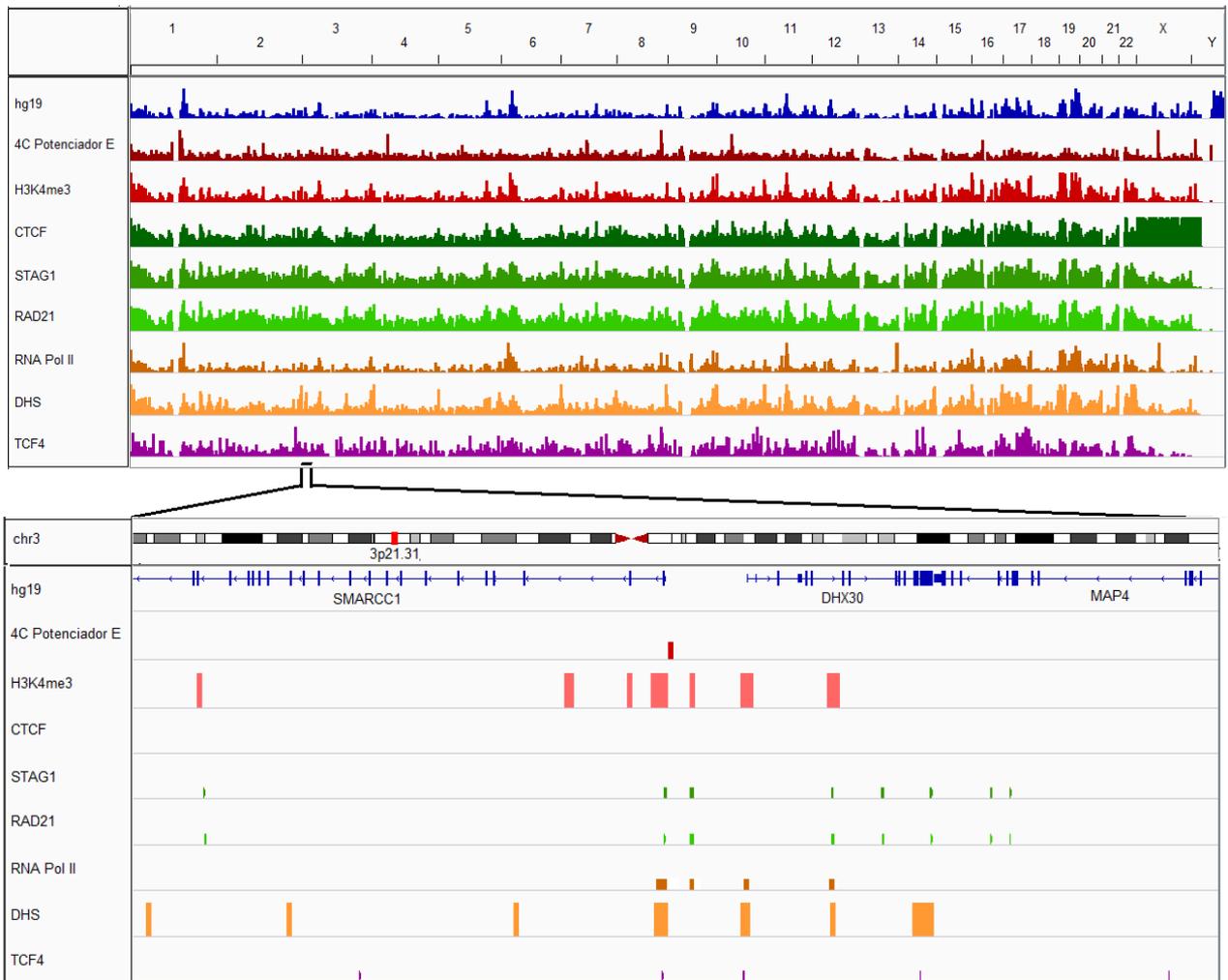


Figura 28 – Colocalización de un pico 4C del potenciador E y marcadores funcionales. Arriba: Panorámica de la distribución de los picos 4C en todo el genoma, acompañado de datos ChIP y DNAsa-Seq. Abajo: Magnificación la interacción SMARCC1 - potenciador E y solapamiento con datos ChIP y DNAsa-Seq.

### Con Circos

Para mostrar los contactos 4C identificados entre cada carnada (potenciador E y el promotor de c-Myc) y sus posibles interactores distribuidos en el genoma (genes y y potenciadores respectivamente), se utilizó el programa Circos. Los datos 4C están acompañados de datos de ChIP y DNAsa-Seq utilizados en esta tesis (Figuras 29 y 30).

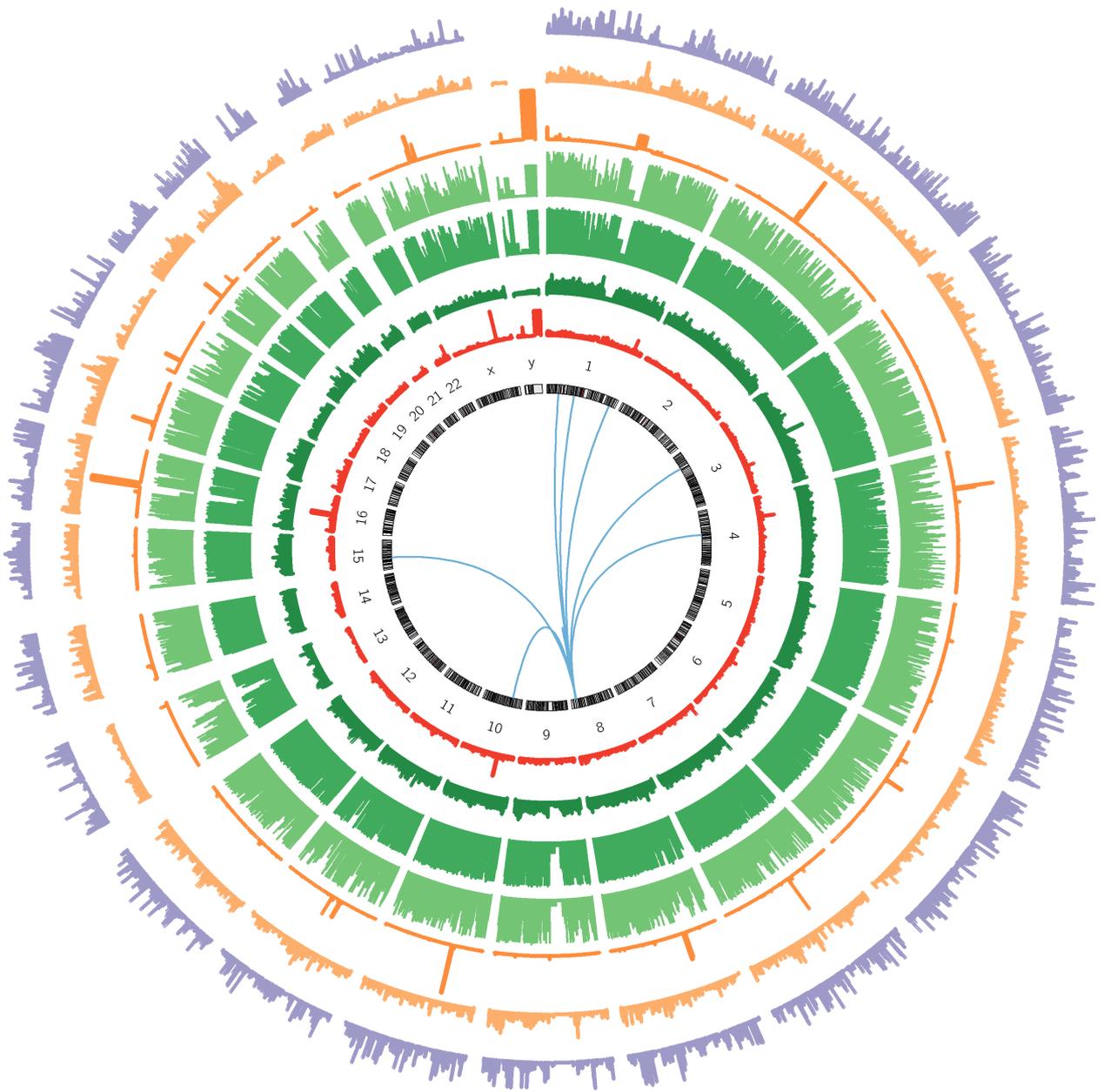


Figura 29 – Contactos genómicos del potenciador E. En esta imagen se aprecian los contactos que establece el potenciador E en 8q24 con sus posibles interactores distribuidos en el genoma. Las líneas color celeste en el centro de la imagen representa los contactos de la carnada con otros puntos genómicos, y en forma circular concéntrica se disponen los datos ChIP y DNase-Seq utilizados para marcar funcionalidad de las regiones. El orden desde adentro hacia afuera: H3K4me3 (rojo), CTCF (verde oscuro), STAG1 (verde), RAD21 (verde claro), RNA Pol II (naranja), DHS (naranja claro), y TCF4 (violeta).

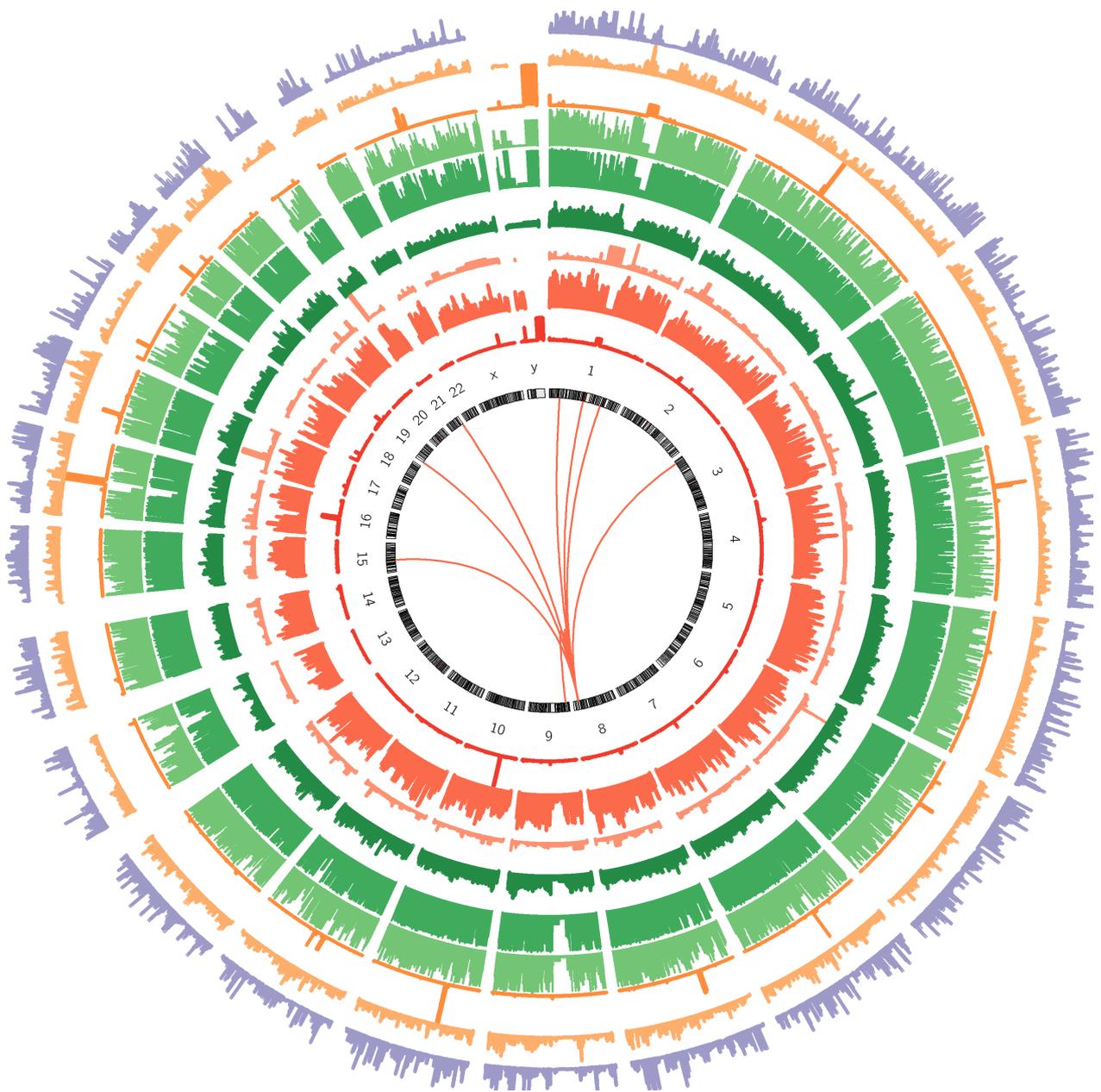


Figura 30 – Contactos genómicos del promotor de c-Myc. En esta imagen se ilustra los contactos que establece el promotor de c-Myc con sus posibles interactores distribuidos en el genoma. Las líneas color rojo en el centro de la imagen representa los contactos de la carnada con otros puntos genómicos, y en forma circular concéntrica se disponen los datos ChIP y Dnasa-Seq utilizados para marcar funcionalidad de las regiones. El orden desde adentro hacia afuera: H3K4me1 (rojo fuerte), H3K27ac (rojo), p300 (rojo suave), CTCF (verde oscuro), STAG1 (verde), RAD21 (verde claro), RNA Pol II (naranja), DHS (naranja claro), y TCF4 (violeta).

### Validación de contactos por 3C-PCR

Para confirmar los contactos entre cada carnada (potenciador E o promotor de c-Myc) y su interactor candidato identificado por 4C (Tablas 14 y 17), se empleó la técnica 3C combinada con PCR con cebadores específicos para cada carnada 4C y su posible interactor que flanquearan el sitio de restricción/ligación (Figura 11, Tabla 4).

Para CCNY y ZNF643, genes interactores del potenciador E, no se pudo diseñar cebadores, debido a que los mismos resultaban ser secuencias repetidas tanto distribuidas por todo el genoma o dentro de la región a estudiar.

De los 5 genes candidatos a interactores del potenciador E, a los cuales se les pudo diseñar cebadores, sólo dos presentaron amplicones del tamaño esperado: ACBD3 y SMARCC1 cuando se corrieron en una electroforesis en gel de acrilamida (Figura 31). Aunque las bandas de estos amplicones son de intensidad débil, se observan pocas bandas correspondientes a amplificaciones inespecíficas. Los otros tres genes que no presentaron el amplicón esperado, también mostraron amplificaciones inespecíficas.

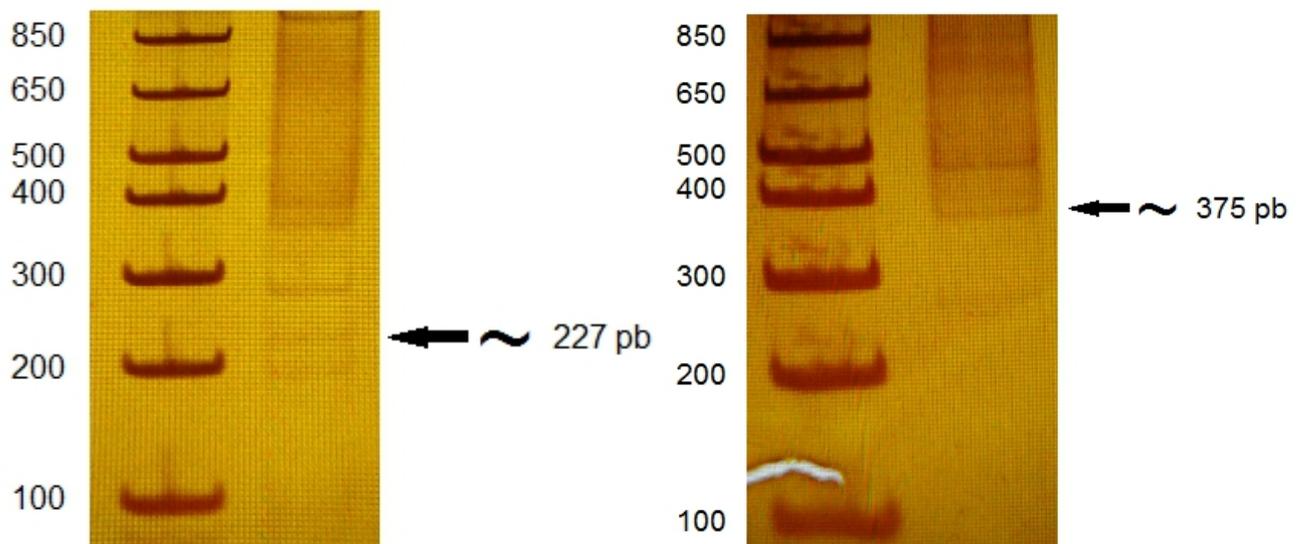
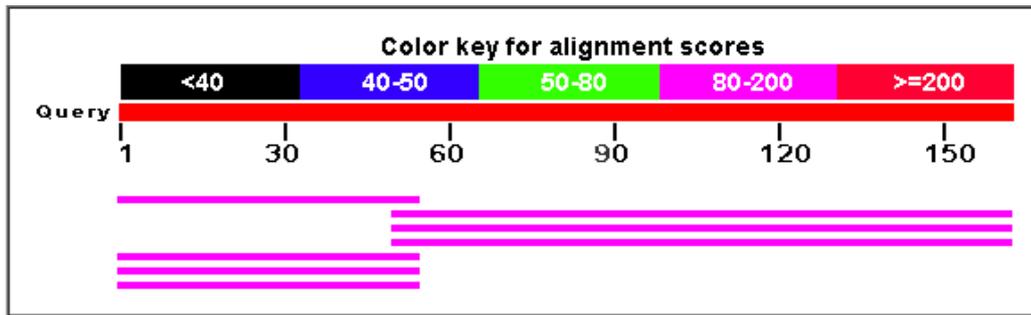


Figura 31 – Confirmación de contactos 4C del potenciador E a través de 3C-PCR. Electroforesis en gel de acrilamida 6% de los contactos del potenciador E con ACBD3 (izquierda) y SMARCC1 (derecha), evidenciados por 3C-PCR con cebadores específicos para los interactores. Las flechas indican la banda correspondiente al amplicón del tamaño esperada para cada caso.

Para corroborar la existencia de las lecturas quiméricas amplificadas por 3C, se buscaron lecturas en los datos 4C que contuvieran fracciones del potenciador E y los interactores ACBD3 y SMARCC1. Para ambos genes se encontraron lecturas quiméricas 4C que al ser alineadas contra el genoma humano a través de BLAST, encontró homología con los cromosomas donde se localizan (Figuras 32 y 33).

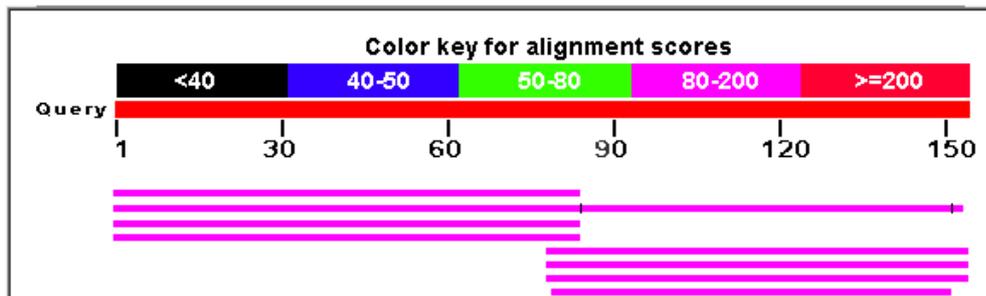
GCAAACCTTTA GCATACCTAG CAATTTAATT GGTGTTCAAG AAATGTCTGA TGATCCTGGT  
 TTTGCTGGAA GCGCAGCTCA TACAATTTGT GGGGACCCTC ATTAGGAAAG AGTACAAAAG  
 TTACGCGCAG AAGGGTAGGA AAGTGAATGT ATATTTAGAA AGA



	Description	Max score	Total score	Query cover	E value	Ident	Accession
<b>Genomic sequences</b>							
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 1, alternate assembly CHM1_1.1</a>	198	198	69%	2e-48	98%	<a href="#">NC_018912.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 1, alternate assembly HuRef</a>	198	198	69%	2e-48	98%	<a href="#">AC_000133.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 1, GRCh38 Primary Assembly</a>	198	198	69%	2e-48	98%	<a href="#">NC_000001.11</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly CHM1_1.1</a>	102	102	33%	2e-19	100%	<a href="#">NC_018919.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly HuRef</a>	102	102	33%	2e-19	100%	<a href="#">AC_000140.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, GRCh38 Primary Assembly</a>	102	102	33%	2e-19	100%	<a href="#">NC_000008.11</a>

Figura 32 – Lecturas quiméricas 4C que contienen al gen ACBD3 y al potenciador E. Arriba: secuencia de lectura quimérica obtenida por 4C, donde se indica en color rojo el sitio de corte de Mbo I que generaría un fragmento de aproximadamente 50 pb y otro de 100 pb. Abajo: Resultados del BLAST contra el genoma humano, donde el largo de cada una de las barras de colores representa el tamaño de cada fracción. Se puede apreciar que los tamaños de las barras son similares a los esperados por la digestión de la lectura. También se muestran los sitios de homología en el genoma (matches), localizados en los cromosomas 1 y 8.

CCTCCCTCTT GGTGTCTATA CCAGTTTTGC AACTTTTAGC ATACCTAGCA ATTTAATTGG  
 TGTTCAAGAA ATGTCTGATG ATCACGAGGT CAGGAGTTCC AGACCACCCT GGCCAACATG  
 GTGAAACCCC GTCTCTACTA AACTACAAA ATTG



	Description	Max score	Total score	Query cover	E value	Ident	Accession
<b>Genomic sequences</b>							
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly CHM1_1.1</a>	156	1.719e+05	100%	1e-35	100%	<a href="#">NC_018919.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly HuRef</a>	156	1.683e+05	100%	1e-35	100%	<a href="#">AC_000140.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, GRCh38 Primary Assembly</a>	156	1.719e+05	100%	1e-35	100%	<a href="#">NC_000008.11</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 3, alternate assembly CHM1_1.1</a>	135	2.554e+05	55%	1e-29	99%	<a href="#">NC_018914.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 3, alternate assembly HuRef</a>	135	2.493e+05	55%	1e-29	99%	<a href="#">AC_000135.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 3, GRCh38 Primary Assembly</a>	135	2.540e+05	55%	1e-29	99%	<a href="#">NC_000003.12</a>

Figura 33 – Lecturas quiméricas 4C que contienen al gen SMARCC1 y al potenciador E. Arriba: secuencia de lectura quimérica obtenida por 4C, en color rojo el sitio de corte de Mbo I que generaría dos fragmentos de aproximadamente 80 pb cada uno. Abajo: Alineamiento contra el genoma humano, los tamaños de las barras (largo de secuencia) son similares a los esperados por la digestión de la lectura. También se muestra los sitios de homología en el genoma, localizados en los cromosomas 3 y 8.

De los 8 elementos potenciadores candidatos a ser interactores de c-Myc, sólo dos presentaron amplicones del tamaño esperado cuando se corrieron en una electroforesis en gel de acrilamida: 1p34.3 y 9p23 (Figura 34). Al igual que para los genes candidatos, las bandas de tales amplicones son intensas pero coexisten bandas de tamaños diferentes.

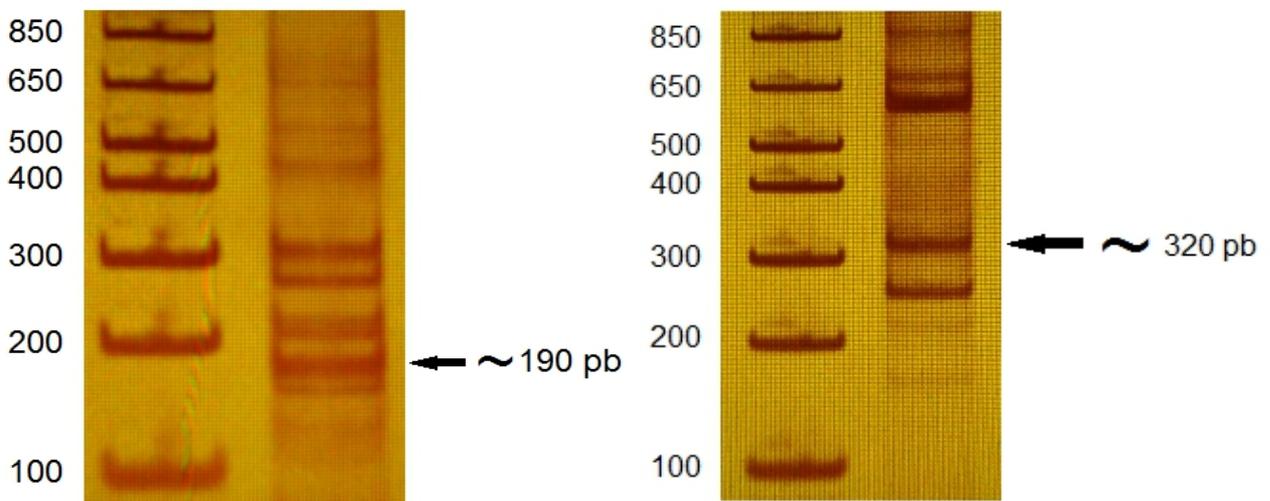
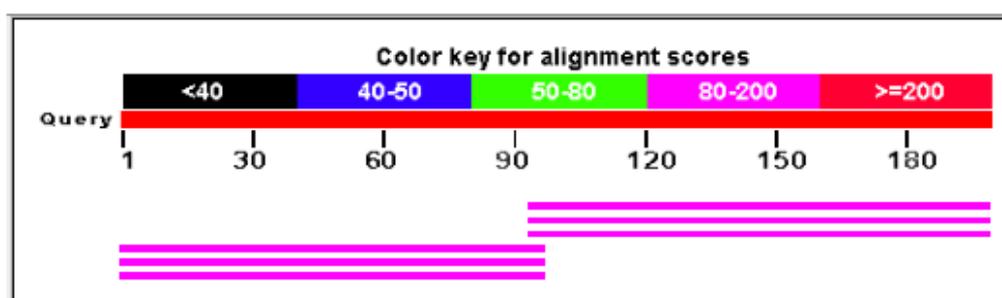


Figura 34 – Confirmación de contactos 4C de c-Myc a través de 3C-PCR. Electroforesis en gel de acrilamida 6% de los contactos de c-Myc con 1p34.3 (izquierda) y 9p23 (derecha). Las flechas indican la banda correspondiente al amplicón del tamaño esperada para cada caso.

Para corroborar la existencia de las lecturas quiméricas amplificadas por 3C, se buscaron lecturas en los datos 4C que contuvieran fracciones de c-Myc y los interactores 1p34.3 y 9p23. Para el caso de 1p34.3, no se encontró lecturas 4C quiméricas que contuvieran a c-Myc y el interactor; mientras que para el caso de 9p23, se encontraron lecturas quiméricas 4C que al ser alineadas contra el genoma humano a través de BLAST, se encontró homología con 9p23 y c-Myc (Figura 35).

```

GGCTTTTCAC CTGAAGCATT CCAAACCAAA ACAGTGAGTG GAAATGTTGT ACAGTGTCTT
CTCAGGCTGC AAAAAGATGA TAATATTTTT ATGGATCAGA ATCGATGCAT TTTTGTGCA
TGACCGCATT TCCAATAATA AAAGGGGAAA GAGGACCTGG AAAGGAATTA AACGTCCGGT
TTGTCCGGGG AGGAAAGAG
  
```



	Description	Max score	Total score	Query cover	E value	Ident	Accession
<b>Genomic sequences</b>							
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly CHM1 1.1</a>	196	196	53%	9e-48	100%	<a href="#">NC_018919.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, alternate assembly HuRef</a>	196	196	53%	9e-48	100%	<a href="#">AC_000140.1</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 8, GRCh38 Primary Assembly</a>	196	196	53%	9e-48	100%	<a href="#">NC_000008.11</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 9, GRCh38 Primary Assembly</a>	174	174	48%	4e-41	99%	<a href="#">NC_000009.12</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 9, alternate assembly CHM1 1.1</a>	163	163	48%	9e-38	97%	<a href="#">NC_018920.2</a>
<input type="checkbox"/>	<a href="#">Homo sapiens chromosome 9, alternate assembly HuRef</a>	163	163	48%	9e-38	97%	<a href="#">AC_000141.1</a>

Figura 35 – Lecturas quiméricas 4C conteniendo al potenciador candidato 9p23 y a c-Myc. Arriba: secuencia de lectura quimérica obtenida por 4C, donde se indica en color rojo el sitio de corte de Mbo I que generaría un fragmento de aproximadamente 90 pb y otro de 100 pb. Abajo: Resultados del alineamiento contra el genoma humano. Se puede apreciar que los tamaños de las barras (largo de secuencias alineadas) son similares a los esperados por la digestión de la lectura. También se muestran los sitios de homología en el genoma (matches), localizados en los cromosomas 9 y 8.

Para aquellos interactores que no amplificaron un producto del tamaño esperado y que presentaban amplicones inespecíficos, también se les encontró lecturas 4C quiméricas que contenían secuencias de la carnada y el interactor en cuestión.

## Comparación con otros datos de interacciones en la línea LNCaP

En esta tesis se identificaron potenciales candidatos a interactuar con el potenciador E y con el promotor de c-Myc en la línea celular LNCaP (Tablas 14 y 17 respectivamente). Un reciente estudio publicado por el grupo de Wang (Wang et al. 2014), reportó las interacciones de largo alcance de los tres bloques de 8q24 asociados con cáncer de próstata (Figura 1) (Witte 2007), en LNCaP y otras líneas celulares.

Este grupo desarrolló la metodología 3C-MTS, que consistió básicamente en preparar una biblioteca 3C, seleccionar las moléculas quiméricas que contenían fragmentos de los distintos bloques de 8q24 con sondas específicas marcadas con biotina (Wang et al. 2014) y secuenciarlas. En el análisis de los datos, excluyeron las regiones  $\pm 25$  Kb desde los bloques en 8q24 para no considerar los eventos de autoligación.

A modo comparativo, se intentó encontrar coincidencias en las interacciones de largo alcance del potenciador E identificadas por 4C, ChIP y DNase-Seq, con las reportadas en LNCaP por el grupo de Wang. Ellos encontraron que una gran fracción de las interacciones involucraban a 6 regiones centroméricas: 1p11.2, 6p11.1, 7q11.21, 10q11, 18q11.1 y 19q11 (Figura 36); pero en la presente tesis se encontró un interactivo en 7q11.23 (Tabla 12), muy próximo a 7q11.21. Por este motivo se decidió comparar todos los picos 4C obtenidos con los datos del grupo de Wang, y no sólo los interactores identificados. En esta segunda instancia, se encontraron cuatro coincidencias más: 1p11.2, 10q.11, 18q11.1 y 19q11.

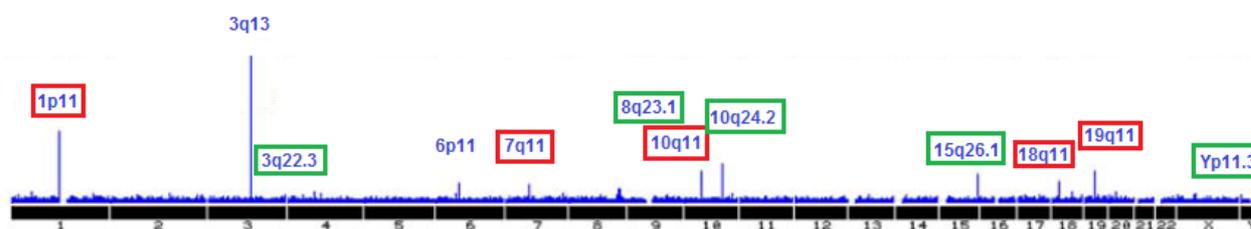


Figura 36 – Comparación de las interacciones de 8q24 en LNCaP en otro estudio de captura. Las interacciones más frecuentes de la región 8q24 de LNCaP observadas por Wang y colaboradores (2014), están representadas por picos de color azul e identificadas con la localización cromosómica. Las regiones marcadas en rojo y en verde corresponden a regiones centroméricas y regiones con mayor densidad de genes respectivamente, que también presentaron picos 4C identificados en esta tesis (Modificado de Wang et al., 2014).

Aparte de los centrómeros, otras interacciones comunes de la región 3 que contenía al potenciador E implicaban a regiones con mayor densidad génica (Tabla 18), las cuales también coincidieron con picos 4C, con la salvedad de que las distancias relativas fueron un poco diferentes a las reportadas por Wang. También se buscó colocalización de las interacciones con datos de ChIP y DNase-Seq utilizados previamente en esta tesis para la caracterización funcional de los interactores (Tabla 18), pero la mayoría no mostró colocalización.

**Tabla 18 – Comparación de las interacciones de 8q24 con otro estudio de captura  
Región 3 conteniendo al potenciador E: 3C-MST vs 4C**

	<b>Wang et al. 2014 (3C-MST)</b>		<b>Tesis (4C)</b>	
	<b>Gen más cercano y posición relativa de la interacción 3C-MTS</b>	<b>Colocalización con CHIP/DNAsa-Seq</b>	<b>Pico 4C del potenciador E más cercanos</b>	<b>Colocalización con CHIP/DNAsa-Seq</b>
3q22.3	UTR 3' de MSL2	DHS (probable)	650 kb río arriba UTR 3' de MSL2	RAD21, DHS
8q23.1	95 kb río abajo del extremo 3' ANGPT1	No	770 kb río abajo del extremo 3' ANGPT1	No
10q24.2	3.3 kb río arriba de RRP12	No	30 kb río abajo de RRP12	No
15.26.1	47 kb río arriba de SLCO3A1	No	900 kb río arriba de SLCO3A1	No
Yp11.3	16.9 kb río arriba de SHOX	No	1.2 Mb río abajo de SHOX	No

## VII. DISCUSIÓN

### **Producción de bibliotecas 4C conteniendo contactos genómicos para el potenciador E y c-Myc en LNCaP**

Para el estudio de las interacciones a nivel genómico de cada región de interés o carnada, se utilizó la metodología 4C (Circular Chromosome Conformation Capture) para construir una biblioteca de moléculas quiméricas conteniendo la secuencia de cada carnada y sus potenciales interactores, finalmente identificados por secuenciación masiva (Figuras 4 y5).

El procedimiento 4C tuvo dos puntos de control: la eficiencia de la digestión y el enriquecimiento por PCR anidado e invertido (Göndör et al. 2008).

Para determinar cualitativamente la eficiencia de la digestión del experimento 4C en esta tesis, se comparó visualmente la electroforesis en gel de agarosa de los tres viales digeridos con las realizadas por Göndör et al. (2008) (Figura 12). Al observar patrones de digestión similares, se procedió a la determinación cuantitativa por qPCR. La eficiencia de la digestión del ADN alcanzó el 78%. De acuerdo con trabajos previos (Göndör et al. 2008), el mínimo aproximado de eficiencia se sitúa en 85%, dado que podría dar lugar a formación de fragmentos de ligación mayores, difíciles de amplificar y detectar. Dos explicaciones a este resultado serían: a) un inexacto conteo de células al comienzo del protocolo, lo que implica un exceso de ADN que no se va a digerir adecuadamente en las condiciones realizadas; b) la formación de agregados durante la preparación para la digestión que dificulta parcialmente la accesibilidad de la enzima de restricción (Göndör et al. 2008).

El enriquecimiento de la carnada y sus posibles interactores contenidos en moléculas circulares se realizó a través de un PCR anidado e invertido de la muestra 4C y los controles ligado y no ligado. Como se esperaba, en el control no ligado no se verificò amplificación, mientras que en la muestra 4C se observó un amplicón correspondiente al evento de autoligación de la carnada, acompañado de otras bandas de mayor tamaño superpuestas en un smear (Figura 13).

La presencia de varias bandas y un smear representa la amplificación de las distintas secuencias interactoras insertas en una molécula circular que contenían a una carnada (Figura 13) (Göndör et al., 2008). Para el control ligado se esperaba que sólo amplificara el producto de la autoligación de la carnada; sin embargo, se observó además un smear similar al de la muestra 4C pero de menor intensidad (Figura 13). Esto podría deberse a que la reversión del entrecruzamiento fue ineficiente y algunas de las moléculas todavía unidas por complejos,

estaban lo suficientemente cercanas para ligarse. No obstante, como es la muestra 4C quien permite identificar los contactos, y el enriquecimiento de cada carnada en la biblioteca 4C mostró el perfil esperado (Figura 13) según Göndör et al. (2008), se puede afirmar que se obtuvo una buena biblioteca 4C de los contactos genómicos del potenciador y el promotor de c-Myc en LNCaP.

### **La identificación de lecturas quiméricas con sitios de restricción para Mbo I indican las regiones ligadas a 8q24 en LNCaP**

La secuenciación de las bibliotecas enriquecidas de la muestra 4C del potenciador E y del promotor de c-Myc, permitió identificar a las regiones que estaban en contacto con cada carnada y posiblemente interactuando. Debido a que el smear del PCR anidado e invertido iba de 150-200 a 1500 pb (Figura 13), y que el fragmento secuenciable es de 400 pb, se fragmentaron los amplicones para un tamaño entre 150-200 hasta 400 pb según indicaciones del fabricante. No se seleccionó por tamaño para no perder fragmentos menores de 400 que podrían ser informativos.

El chip se cargó un 46%, lo que permitió obtener más de 3 millones de lecturas de un promedio de 150 pb entre las dos librerías (Potenciador E y promotor de c-Myc) (Figura 17). El porcentaje de enriquecimiento fue satisfactorio (99%), indicando que solo el 1% de los pocillos cargados en el chip tenían ISPs sin molde. Los porcentajes de ISPs policlonales (16%) y de baja calidad (27%) estuvieron dentro de las expectativas indicadas por el fabricante; y no se detectó dímeros de cebadores de secuenciación.

El largo promedio de longitud de las lecturas fue menor al esperado en la secuenciación: 153 pb para el Potenciador E, 143 pb para c-Myc (Tabla 5). Aunque se trató de optimizar el tiempo de fragmentación de los amplicones dentro del rango propuesto por el fabricante, no se logró un promedio mayor (Figuras 14 y 15), pero sí se registraron moléculas de 400 pb.

Con la remoción de nucleótidos de baja calidad y el filtrado por el largo de las lecturas obtenidas (Tabla 6, Figura 18), se obtuvieron 1495340 lecturas para el potenciador E y 1021207 para el promotor de c-Myc, números que se encuentran dentro de lo aceptado para un experimento 4C (De Laat et al. 2012).

Al alinear las lecturas de cada experimento 4C al genoma de referencia, se obtuvo un 99.7% de alineamiento cuando el potenciador E fue la carnada, y 97.4% cuando lo fue el promotor de c-

Myc (Tabla 6). Más del 60% de las lecturas alinearon dentro de la región situada a  $\pm 2$  kb de cada carnada, valor aceptado para una experimento de buena calidad (De Laat et al. 2012).

El 20.5% de las lecturas alineadas para cada carnada correspondió a lecturas quiméricas (Tabla 7), que contenían secuencias provenientes de hasta 6 regiones genómicas diferentes (Tabla 8); mientras que un promedio del 7% eran quimeras conteniendo la región 8q24 (Tabla 7). La búsqueda de homología en el genoma humano de algunas de estas lecturas a través del BLAST, y la verificación de sitios de reconocimiento para Mbo I demostraron su naturaleza quimérica (Figuras 19 y 20).

Estas lecturas quiméricas surgieron de la secuenciación de fragmentos de los productos del PCR invertido y anidado, fragmentos que contenían al punto de ligación entre la carnada y su interactor. Esto las hace muy útiles para demostrar que dos regiones del genoma realmente se ligaron. Probablemente las lecturas conteniendo 2 o hasta 3 fracciones representen eventos de interacción entre varias regiones, y las que contenían de 4 a 6 resultan mas difíciles de interpretar.

En suma, la secuenciación de lecturas quiméricas conteniendo diferentes sitios genómicos, y la presencia de los correspondientes sitios de restricción de Mbo I en las uniones, indican que se capturaron los contactos genómicos en la línea LNCaP para los que estos experimentos fueron diseñados.

#### **Picos 4C se corresponderían con los amplicones producidos por el PCR invertido y anidado**

Para encontrar alguna correlación entre los productos del PCR invertido y anidado y los picos 4C, se comparó el rango de los productos de PCR (smear) con la extensión de los picos. El rango del smear del PCR era de 150 – 200 hasta 1500 pb para las dos carnadas (Figura 13); mientras que el rango de extensión de los picos 4C era de 137 a 3152 pb para el potenciador E, y de 202 a 3857 pb para c-Myc. La superposición de los rangos de tamaño de los amplicones y la extensión de los picos 4C identificados, sugiere claramente que los picos 4C corresponderían a los productos de PCR invertido y anidado.

## **Los genes cercanos a los picos 4C compartidos entre el potenciador E y c-Myc tienen funciones relacionadas con desarrollo, muerte y sobrevivencia celular.**

El estudio ontológico de los genes más cercanos a los picos 4C en común a ambas carnadas, reveló un gran número de genes vinculados con cáncer (Tabla 9) y con funciones relacionadas con desarrollo, crecimiento, proliferación, muerte y sobrevivencia celular. El estudio sobre los genes más cercanos a los picos 4C más abundantes para cada carnada, mostró que poseían funciones relacionadas con desarrollo y movimiento celular, pero aparentemente no están asociados con cáncer. Esto puede deberse a que dichos genes no tengan vínculo establecido o inferido con cáncer, o que sí lo tengan regiones desconocidas, no anotadas aún en el genoma humano.

## **El mayor porcentaje de picos 4C de ambas carnadas se localizaron en el cromosoma 8**

Para estudiar la distribución de los picos 4C del potenciador E y del promotor de c-Myc en el genoma se empleó el programa CEAS, quien compara la distribución de los picos con regiones mapeables del genoma.

Se observó que tanto el potenciador E como c-Myc presentan el mayor porcentaje de sus picos en el cromosoma 8 donde se localizan (7.1% y 7.3% respectivamente) en relación al background (5.1%) (Figura 21), coherente con lo obtenido por Wang y colaboradores (Wang et al. 2014a), quienes estudiaron las interacciones de la región 8q24 en varias líneas celulares, incluyendo a LNCaP.

Este elevado porcentaje de picos 4C en el cromosoma 8 puede explicarse porque la mayoría de los mismos corresponden a la autoligación de los fragmentos de restricción que contenían a cada carnada o a ligaciones de la región 8q24 con regiones cercanas en el mismo cromosoma, que tuvieron lugar por la cercanía y no necesariamente por ser una interacción (de Laat and Dekker 2012). Cabe aclarar que este porcentaje también incluye las interacciones locales previamente descritas, tales como el potenciador E y c-Myc (Pomerantz et al. 2009; Sotelo et al. 2010); y que se observó una interacción cruzada entre ellos.

## Regiones río abajo de genes que interactúan con el promotor de c-Myc podrían constituir potenciadores múltiples o “shadow enhancers”

Al igual que para el genoma, también se estudió la distribución y posicionamiento de los picos 4C respecto a las distintas características genómicas: promotores, UTRs, etc. a través de CEAS. Aproximadamente el 40% de los picos 4C de cada carnada se localizó en intrones, y más del 50% en regiones intergénicas distales (Figura 22). Para el caso de los picos 4C de c-Myc, que representarían elementos potenciadores, concuerda con los resultados obtenidos por Tan y colaboradores (Tan et al. 2014), quienes encontraron que la mayoría de los potenciadores se localizaban en intrones, seguido por regiones intergénicas y UTR 3' (Figura 37).

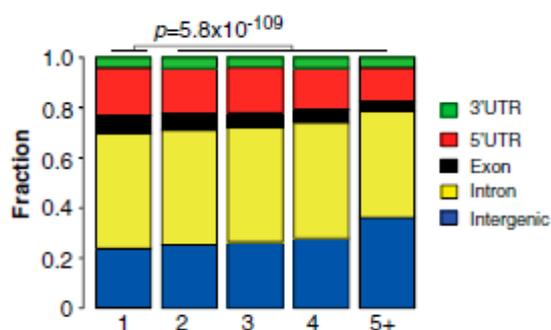


Figura 37 – Distribución genómica de elementos potenciadores en otras líneas celulares. En el eje de las X se indica el número de potenciadores que controlan a un promotor y en el de las Y la fracción de promotores que son controlados por los potenciadores (Extraído de Tan et al. 2014).

El grupo de Tan (Tan et al. 2014) estudió la distribución genómica de múltiples potenciadores que controlaban a un mismo promotor, conocidos como “shadow enhancers” (Hong et al. 2008). Ellos encontraron que, a medida que un promotor era controlado por más potenciadores, éstos se localizaban mayoritariamente en regiones intergénicas (Figura 37). Otro trabajo reciente también reportó ejemplos de promotores contactando múltiples potenciadores (Sanyal et al. 2012).

En esta tesis se observó un enriquecimiento de picos 4C de c-Myc en regiones río abajo de genes, que se incrementaba a medida que lo hacía su distancia al TTS (Figura 24A). En caso de que estos picos correspondan a elementos potenciadores (según criterios funcionales que se aplicarán más adelante), y se demuestre que interactúan con el promotor de c-Myc (a través de c-Myc), surge la posibilidad de considerarlos como “shadow enhancers”: múltiples potenciadores que controlarían a c-Myc.

## **Caracterización funcional de interactores por características de su cromatina**

Para identificar genes candidatos a ser controlados por el potenciador E, o elementos potenciadores candidatos a regular la actividad de c-Myc, se cruzaron los picos 4C de cada carnada con datos provenientes de CHIP y DNAsa-Seq (Tabla 1). Estos datos corresponden a características de la biología de la cromatina específica promotores y potenciadores involucrados en interacciones de largo alcance (Sanyal et al. 2012; Tan et al. 2014).

### Estudio de marcadores de promotores y potenciadores

Como marcador específico de regiones promotoras se utilizó a H3K4me3 (trimetilación del residuo de lisina 4 de la histona 3), como marcadores específicos de elementos potenciadores se utilizó a H3K4me1 (monometilación), H3K27ac (acetilación del residuo de lisina 27 de H3) y la proteína p300 (Visel et al. 2009; Handoko et al. 2011).

### *22 genes serían candidatos a interactuar con el potenciador E*

De la intersección de sitios con modificaciones tipo H3K4me3 con los picos 4C del potenciador E, se encontraron 156 regiones que presentaban ambas características, y 178 genes asociados a ellas (Figura 25). De estos 178 genes, 22 presentaban la intersección dentro de su promotor, definido entre -2000 y +500 pb respecto al TSS (Tabla 11).

### *94 potenciadores serían candidatos a interactuar con c-Myc*

Cuando se interseccionaron los picos 4C de c-Myc con las tres características de potenciadores seleccionadas (H3K4me1, H3K27ac, p300), no hubo resultados, por lo que se procedió a interseccionar picos 4C con dos de los tres marcadores. Se encontraron 86 regiones que solapaban picos 4C con H3K4me1 y H3K27ac (Figura 26, Tabla 15); y 8 regiones que solapaban 4C con p300 (Tabla 16), de las cuales 7 que coincidían además con H3K4me1 (Figura 38). La falta de coincidencia de los tres marcadores para elementos potenciadores podría deberse al bajo número de picos contenidos en los datos de CHIP-Seq de p300 (902 picos, contra 69585 de H3K4me1 y 25000 de H3K27ac), por lo que se podría estar subestimando la cantidad.

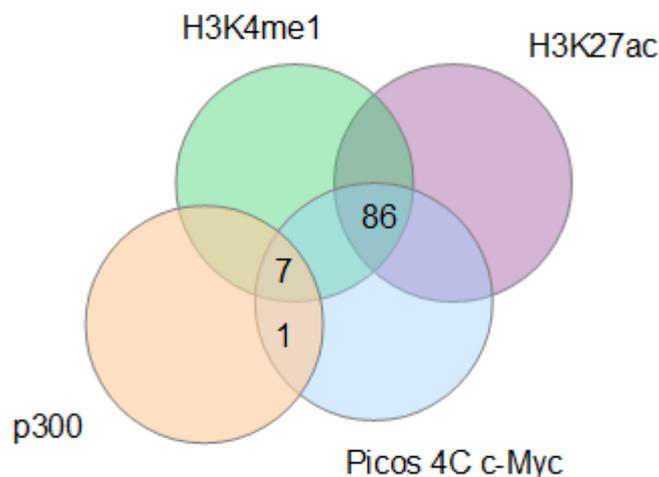


Figura 38 – Marcadores de elementos potenciadores en picos 4C de c-Myc, H3K4me1, H3K27ac y p300.

Sin embargo, un reciente estudio de captura de conformación de la cromatina realizado por Tan y colaboradores (ChIA-PET) (Tan et al. 2014), mostró una gran variación en el porcentaje de potenciadores predichos por H3K4me1 y H3K27ac, que contenían un sitio para p300 en varias líneas celulares, que iba desde 8 a 71% (Tan et al. 2014), lo que contempla al pequeño porcentaje (9%) de candidatos a potenciadores seleccionados por p300 (8 potenciadores predichos por p300 en 92 identificados en total).

### **Genes y potenciadores candidatos (predichos por p300 y H3K4me1) muestran características de cromatina abierta, sugiriendo actividad transcripcional**

Para que los elementos candidatos estén interactuando deben estar funcionalmente activos, deben estar transcribiéndose o regulando a alguien. Esta actividad implica tener una conformación de cromatina “abierta”, accesible para la maquinaria transcripcional y a la acción de DNAsas. Como indicadores de accesibilidad, se buscaron sitios de unión para la RNA Pol II y sitios de hipersensibilidad a la DNasa (DHS) en los candidatos (Kim et al. 2010).

De los 22 genes candidatos, 77% tenía sitios para RNA Pol II; 63% eran DHS y 50% eran DHS con sitios para RNA Pol II (Tabla 13, resumen en Tabla 19). De los 86 potenciadores candidatos predichos por modificaciones de histonas (H3K4me1 y H3K27ac) 33% tenían sitios para RNA Pol II, 63% eran DHS y 14% eran DHS con sitios para RNA Pol II (Tabla 15, resumen en Tabla 19); mientras que de los 8 candidatos predichos por p300, 88% tenían sitios para RNA Pol II, 63% eran DHS y 50% eran DHS con sitios para RNA Pol II (Tabla 16, resumen en Tabla 19).

<b>Tabla 19 – Marcadores de cromatina abierta en genes y potenciadores candidatos</b>			
	RNA Pol II (%)	DHS (%)	RNA Pol II y DHS (%)
Genes predichos por H3K4me3	77	63	50
Potenciadores predichos por p300	88	63	50
Potenciadores predichos por H3K4me1 y H3K27ac	33	63	14

El porcentaje observado para DHS es apoyado por los datos generados por Tan y colaboradores (Tan et al. 2014), quienes encontraron que en varias líneas celulares más del 50% de los potenciadores predichos eran DHS. La utilización de datos de CHIP para RNA Pol II de la línea celular VCaP, derivada de una metástasis vertebral de cáncer de próstata, podría causar la menor colocación de RNA Pol II y DHS en potenciadores predichos por modificaciones de histonas, sugiriendo que los mismos no se están transcribiendo en VCaP. No obstante, no se encontró este problema en la predicción de potenciadores activos con p300, ya que un 88% de los mismos tenían sitios para RNA Pol II, e incluso un 75% contenía al marcador H3K4me1 (Tabla 16). Esto sugiere que para la predicción de potenciadores activos, la sola combinación de p300 con H3K4me1 sería más adecuada que incluir a H3K27ac.

### **Las cohesinas STAG1 y RAD21 mediarían, más que CTCF, los lazos de cromatina implicados en los contactos 4C identificados**

Si dos elementos distales están estableciendo un contacto, éste puede ser mediado por lazos de cromatina que acerque a los elementos. Los sitios de unión para proteínas implicadas en la formación de dichos lazos, tales como CTCF y las cohesinas RAD21 y STAG1, se emplearon para definir cuáles genes o potenciadores candidatos están interactuando con cada una de las carnadas. Recientemente se demostró el rol de CTCF en la interacción entre el potenciador E y c-Myc, a través del knockdown de CCAT1-L que redujo la unión de CTCF al ADN así como la frecuencia de la interacción, evidenciado por CHIP y 3C en líneas celulares de CRC (Xiang et al. 2014).

El laboratorio del grupo de Dekker (Sanyal et al. 2012) estudiaron interacciones de largo alcance en líneas celulares de leucemia mieloide crónica (LMC), de linfopatías y de tumor de cuello uterino a través de captura de conformación de la cromatina (5C). Ellos encontraron que el 79% de las interacciones contenían al menos un sitio de unión para CTCF (Sanyal et al., 2012). En

esta tesis se encontró que el 32% de los 22 genes candidatos tenían un sitio de unión para CTCF (Tabla 13, resumen Tabla 20), mientras que para los potenciadores candidatos predichos por modificaciones de histonas y por p300 el porcentaje fue menor, 26 y 12% respectivamente (Tablas 15 y 16, resumen Tabla 20). Para explicar estos bajos porcentajes en los potenciadores, cabe recordar que los mismos fueron seleccionados en base a datos de RNA Pol II de VCaP y que probablemente no se están transcribiendo, interactuando o formando lazos de cromatina mediados por CTCF o cohesinas.

<b>Tabla 20 – Marcadores de lazos de cromatina en genes y potenciadores candidatos</b>			
	CTCF (%)	RAD21 (%)	STAG1 (%)
Genes predichos por H3K4me3	32	63	77
Potenciadores predichos por p300	12	22	524
Potenciadores predichos por H3K4me1 y H3K27ac	26	25	25

El mismo grupo también observó que el 58% de las interacciones colocalizaban con CTCF y RAD21 (Sanyal et al., 2012), mientras que en esta tesis se observó para el 27% de los genes candidatos (Tabla 13, resumen Tabla 20), 6% para potenciadores predichos por modificaciones de histonas (Tabla 15, resumen Tabla 20) y 25% para los predichos por p300 (Tabla 16, resumen Tabla 20). Esta diferencia puede residir en que los datos de ChIP para RAD21 (y también STAG1) provenían de MCF-7, una línea celular derivada de cáncer mamario (Tabla 1).

Inicialmente en esta tesis se incluyó una ChIP para CTCF de MCF-7; al comparar LNCaP con MCF-7 se encontró que éste último tenía menor número de picos y que el 59% de los sitios de ambas líneas celulares colocalizaban. Esto motivó a utilizar datos de ChIP para cohesinas de MCF-7 ante la falta de datos de LNCaP, pero los resultados aquí presentados sugieren que su uso probablemente subestimó el número de sitios de unión de cohesinas para LNCaP y no consideró a otros.

Sin embargo, el mayor porcentaje de sitios de cohesinas en los elementos candidatos comparado con CTCF, tanto en genes como potenciadores candidatos predichos por histonas (Tablas 13, 15 y 16), se observó en otras líneas celulares estudiadas por el grupo de Tan (Tan et al. 2014) (carcinoma hepatocelular, tumor de cuello uterino, LMC y linfoblastoide). En el caso de los potenciadores predichos por p300, hay una coincidencia del 100% de sitios que colocalizan CTCF y cohesinas (Tabla 16). Se ha propuesto que las cohesinas facilitarían la formación de lazos de cromatina mediados por CTCF (Dorsett 2011); pero la identificación de sitios de unión

independientes de CTCF (Faure et al. 2012) y su interacción con factores de transcripción específicos de tejidos (Tan et al. 2014) sugieren otro papel en torno a los lazos. Podrían mediar la formación de los lazos sin CTCF (Tan et al. 2014) o mantener el lazo inicialmente establecido a través de la unión de reguladores maestros de la transcripción (Schmidt et al. 2010). Por este motivo no sorprende la mayor cantidad de sitios de unión para cohesinas en los elementos candidatos, ya que podrían estar mediando interacciones independientemente de CTCF.

### **Los sitios de TCF4 involucrados en los contactos 4C presentan características de cromatina abierta**

Por último, se incluyó a TCF4 porque posee múltiples sitios de unión en la región promotora de c-Myc y un sitio en el potenciador E, lo que da la posibilidad de que esté también presente en los candidatos.

El 14% de los genes candidatos tenían sitios de unión para TCF4, y todos ellos coincidían con sitios de unión para cohesinas, RNA Pol II y DHS: c-Myc, PEX1 y RBM48 (éstos dos últimos compartían una misma intersección) (Tabla 13). El 6% de los potenciadores candidatos predichos por modificaciones de histonas tenían sitios para TCF4, y coincidían con sitios para cohesinas, RNA Pol II y DHS (Tabla 15); mientras que 13% de los potenciadores predichos por p300 (1 elemento) tenían sitios para TCF4 y colocalizaban con H3K4me1, RNA Pol II y DHS, pero no con cohesinas ni CTCF (Tabla 16). Estos resultados se compararon con los de un trabajo que reportó que entre 25 y 57% de potenciadores predichos por H3K4me1 y H3K27ac en varias líneas celulares derivadas de cáncer (entre ellas MCF-7), contienen sitios de unión a TCF4 (Fietze et al. 2012). El menor porcentaje observado en esta tesis (6 y 13%) (Tablas 15 y 16), podría deberse a que la predicción de los potenciadores con H3K4me1 y H3K27ac no resultó tan acertada como con p300.

### **Regiones funcionales seleccionadas como candidatos a interactuar con el potenciador E o promotor de c-Myc, presentan cromatina abierta y sitios de unión para cohesinas**

Luego de cruzar picos 4C con datos de ChIP y DNase-Seq, y de revisar la literatura y niveles de expresión génica, se seleccionaron 7 genes y 8 potenciadores como candidatos a establecer interacciones con el potenciador E y el promotor de c-Myc (Tablas 14 y 17).

De los 7 genes identificados, la mayoría presentaban cromatina abierta, indicado por la presencia de sitios para RNA Pol II e hipersensibilidad a la DNAsa I, la mitad presentaba sitios para cohesinas, sólo uno contenía un sitio de unión para CTCF (gen BCL9), y ninguno para TCF4 (Tabla 11). Los que tenían mayor nivel de sobreexpresión en muestras clínicas fueron tres: BCL9, SMARCC1 y ZNF643.

De los 8 potenciadores candidatos, la mayoría presentaban cromatina abierta, marcadores epigenéticos de potenciadores (H3K4me1 y H3K27ac), y sitio de unión para cohesinas (Tabla 15). Menos de la mitad de los potenciadores tenían sitios de unión para CTCF, TCF4 y otro marcador de potenciadores, p300 (Tabla 15).

### **Cada región estudiada puede interactuar con varias regiones genómicas**

La interacción entre cada carnada y su potencial interactor identificado por 4C, se verificó a través de 3C-PCR.

De los 7 genes candidatos a interactuar con el potenciador E, se confirmó la interacción de dos de ellos, ACBD3 y SMARCC1 (Figura 31). El gen ACBD3 codifica para una proteína involucrada en el mantenimiento y función del aparato de Golgi y en el control del ciclo celular, y podría ser regulada por el supresor de tumores RB (ver Tabla 14). El gen SMARCC1 codifica una proteína miembro de la familia SWI/SNF, que regularía la transcripción por alteración de la estructura de la cromatina, y que se sobreexpresa en diversos tejidos tumorales (ver Tabla 14). A su vez, se encontraron lecturas quiméricas 4C que contenían a la secuencia del potenciador y de cada uno de los genes (Figuras 32 y 33), sustentando los resultados del 3C-PCR.

El número total de interactores reportado en esta tesis, concuerda con Tan y colaboradores (Tan et al. 2014), quienes observaron que la mayoría de los potenciadores predichos interactúa con un solo promotor, y fracciones menores lo hacen hasta más de 5, promediando en total 2.92 promotores (Figura 39). La regulación de varios genes por parte de un potenciador, donde colocalicen sitios para RNA Pol II, CTCF y/o cohesinas, podría constituir una fábrica de transcripción donde se coordine la actividad de dichos genes (Ong and Corces 2014; Pombo et al. 1999).

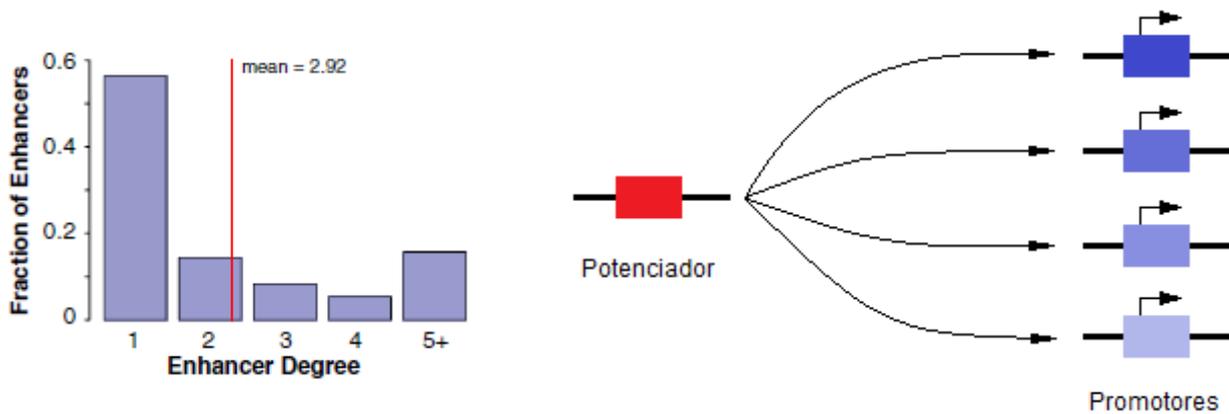


Figura 39 – Potenciadores que interactúan con uno o más promotores. Izquierda: Fracción de potenciadores que regulan a uno o más promotores (Extraído de Tan et al. 2014). La línea vertical roja indica el promedio estimado. Derecha: esquema de un potenciador actuando sobre varios promotores.

De los 8 elementos potenciadores candidatos a regular la actividad de c-Myc (Tabla 17), se confirmó la interacción de dos de ellos a través de 3C-PCR: el elemento localizado en 1p34.3 y en 9p23 (Figura 34). El grupo de Tan también reportó que un promotor interactúa con un promedio 2.02 potenciadores (Figura 40); sugiriendo nuevamente el concepto de “shadow enhancers” (Hong et al. 2008). Otro estudio también reportó un promedio similar, de 1.84 potenciadores por promotor (Sanyal et al. 2012). Por otro lado, al intentar buscar apoyo en los datos 4C, solo se encontró lecturas quiméricas 4C que contenían secuencias del potenciador candidato en 9p23 y c-Myc (Figura 35).

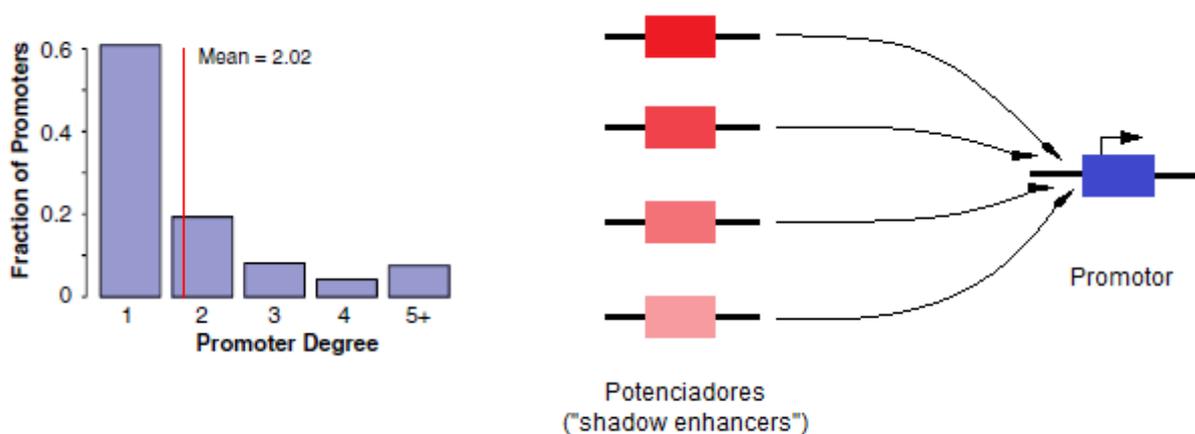


Figura 40 – Promotores que interactúan con uno o más potenciadores. Izquierda: Fracción de promotores que son regulados por uno o más potenciadores (Extraído de Tan et al. 2014). La línea vertical roja indica el promedio estimado. Derecha: esquema de varios potenciadores actuando sobre un mismo promotor.

En la presente tesis se confirmó que el potenciador E puede interactuar con al menos dos genes vinculados a cáncer, ACBD3 y SMARCC1, y que c-Myc lo haría con al menos un elemento potenciador localizado en 9p23. Hasta el momento, no se ha reportado a ACBD3 ni a SMARCC1 como blancos del potenciador E, ni la existencia de un elemento regulador en la región 9p23 cercano al punto de interacción. A estas interacciones hay que sumar la del potenciador E y c-Myc, exhaustivamente estudiada y confirmada por 3C en trabajos previos (Pomerantz et al. 2009; Sotelo et al. 2010).

Sin embargo, la existencia de lecturas quiméricas 4C para varios de los otros candidatos no confirmados por 3C, en algunos casos fuertemente vinculados a cáncer ya sea por la literatura o por su sobreexpresión, sugiere que tanto el potenciador E como el promotor de c-Myc podrían establecer un mayor número de contactos que los confirmados por 3C. Esta sugerencia tiene como sustento el trabajo de Tan que reporta interacciones con hasta más de 5 elementos (Tan et al. 2014), y el de Wang que propone a la región 8q24 como un centralizador de contactos (Wang et al. 2014a).

Resulta interesante pensar en la cantidad de contactos genómicos que establece cada región de estudio, si se da en simultáneo podría constituir una fábrica de transcripción (Pombo et al. 1999), o ser consecuencia de una población celular heterogénea y aneuploide como lo es LNCaP donde coexisten células desincronizadas y con varios alelos de 8q24.

### **Comparación con otras metodologías del tipo 3C en LNCaP mostró algunas coincidencias**

Para verificar las interacciones identificadas en la presente tesis, se compararon los resultados obtenidos con un trabajo publicado recientemente acerca de las interacciones de largo alcance de los tres bloques de la región 8q24 en líneas celulares derivadas de tejido normal y tumoral de próstata, incluida LNCaP (Wang et al. 2014a).

Lo primero que resaltaron los autores del trabajo, es que una gran fracción de las interacciones involucraban a algunas regiones centroméricas (Figura 36). De entre los 22 genes candidatos para el potenciador E, se encontró que uno, el gen NSUN5, localizaba en un centrómero (Tabla 11). Cuando se consideraron todos los picos 4C del potenciador E, se encontró un mayor número de coincidencias. Sin embargo, los autores no indicaron a cuál bloque correspondía cada interacción ni la posición genómica específica de las mismas, lo que hace difícil la evaluación de las coincidencias.

Respecto a las interacciones con regiones de mayor densidad génica, sí se detalló el bloque involucrado y la posición de cada interacción relativa al gen más cercano. Dentro de las interacciones comunes del bloque 3 de 8q24 que contenía al potenciador, se encontraron algunas coincidencias con los picos 4C, aunque las distancias relativas variaban un poco de las reportadas por Wang (Tabla 18). En la gran mayoría de las coincidencias, no se encontró colocalización con marcadores de promotores u otras características funcionales (Tabla 18).

Las diferencias entre esta tesis y el trabajo, pueden deberse a la metodología utilizada y análisis de los datos. El grupo de Wang utilizó la metodología 3C-MTS, una variante de 3C que consistió en preparar una biblioteca 3C, seleccionar las moléculas quiméricas que contenían fragmentos de los distintos bloques de 8q24 con sondas específicas marcadas con biotina y secuenciar (Wang et al. 2014a). En 3C-MTS se utilizó a EcoRI para fragmentar la cromatina entrecruzada, una enzima de restricción de corte raro que genera fragmentos de mayor tamaño que una de corte frecuente como en 4C, por lo que la resolución de la técnica podría ser menor (Göndör et al. 2008). Adicionalmente, respecto al análisis de datos, los autores no utilizaron datos de ChIP o DNAsa-Seq para estudiar la funcionalidad de los contactos, por lo que pueden estar considerando contactos al azar como específicos.

La poca información que brindan los autores sobre las posiciones específicas de los contactos, así como la falta de estudios de colocalización de marcadores para promotores, no permitieron encontrar mayores coincidencias con esta tesis, lo cual hubiese sido interesante para reafirmar los resultados obtenidos.

## VIII. CONCLUSIONES Y PERSPECTIVAS

### Conclusiones

Se identificaron 7 genes candidatos a interactuar con el potenciador E, y 8 posibles elementos potenciadores que regularían a c-Myc en LNCaP. Varios de los genes identificados tienen vinculación funcional con procesos tipo cáncer.

Se confirmaron 3 de estos contactos genómicos: los genes ACBD3 y SMARCC1, vinculados con cáncer, y un posible elemento potenciador en 9p23.

Se identificaron lecturas quiméricas conteniendo sitios de restricción para la enzima Mbo I y homología a las secuencias de las carnadas y sus candidatos correspondientes, evidenciando la existencia de los contactos en la región 8q24 en LNCaP.

Se observó que la mayoría de los genes y potenciadores predichos muestran características de cromatina abierta, sugiriendo actividad transcripcional de los mismos.

Se encontró que las cohesinas STAG1 y RAD21 median la formación de los lazos de ADN implicados en las interacciones de largo alcance en LNCaP, de forma independiente a CTCF.

En la presente tesis se obtuvieron bibliotecas 4C del potenciador E y de c-Myc, dos elementos genómicos localizados en la región 8q24, con suma importancia en cáncer de próstata. La secuenciación de estas bibliotecas permitió identificar los contactos que cada uno de estos elementos establece con el resto del genoma humano.

Para analizar las interacciones identificadas por 4C, se planteó una estrategia que integró información de ensayos de inmunoprecipitación de la cromatina y DNAsa, en su mayoría provenientes de la línea LNCaP. Esto argumentó y direccionó el análisis de esta tesis hacia resultados de depuración funcional de las interacciones, así como también respaldó su comparación con trabajos similares.

## Perspectivas

Para demostrar los roles biológicos de los interactores confirmados por 3C-PCR, y saber si están implicados en una misma red de interacción, realizar ensayos funcionales en vectores de expresión, en presencia o ausencia de ciertos factores. En el caso del posible potenciador 9p23, estudiar la expresión de la zona en tejido normal y tumoral, y correlacionarlo con niveles de c-Myc.

Hubo regiones que no se analizaron con profundidad, por ejemplo picos 4C del potenciador E que no eran próximos a un promotor. Estudios *in silico* en estas regiones para detectar estructuras génicas, o de su expresión en LNCaP, permitirá encontrar nuevos genes que no estaban anotados o interacciones con una función más compleja.

Esta tesis también constituyó una experiencia para mejorar el uso de herramientas informáticas, tales como los comandos en línea en el sistema Linux (grep, awk, sed, sort, uniq y wc), perl y python, y programas específicos para el análisis de las lecturas de secuenciación. Esto motiva fuertemente a desarrollar o mejorar su uso para lograr nuevos objetivos, como puede ser estudiar a nivel masivo la correspondencia de lecturas quiméricas alineadas en distintos sitios genómicos.

Una interrogante que surgió durante esta tesis, fue pensar si el número de interacciones identificadas por metodologías basadas en captura de la conformación de la cromatina, están presentes en una sola célula, o son consecuencia de la sumatoria de las interacciones presentes en la población celular de estudio. La región 8q24 está amplificada en varios tipos de cáncer, por lo que se podría estar capturando las interacciones de cada copia en simultáneo. Lo mismo podría suceder para una línea celular como LNCaP, hipotetraploide, que puede contener varios alelos de la región. La realización de experimentos 4C en cultivos celulares sincronizados o el empleo alguna técnica que permita discriminar entre los alelos o copias de 8q24, son opciones interesantes para responder esta interrogante.

A modo confirmatorio, utilizar los interactores identificados en esta tesis como carnadas en nuevos experimentos 4C, pero también a modo exploratorio para incrementar el número de posibles integrantes de redes de interacción y su co-participación en distintas redes.

Como confirmación independiente a técnicas basadas en 3C, utilizar FISH para verificar contactos físicos encontrados con captura de la conformación de la cromatina (Spilianakis *et al.* 2005; Zhao *et al.* 2006; Lomvardas *et al.* 2006).

## IX. BIBLIOGRAFÍA

- Adachi S, Jigami T, Yasui T, Nakano T, Ohwada S, Omori Y, Sugano S, Ohkawara B, Shibuya H, Nakamura T, et al. 2004. Role of a BCL9-related beta-catenin-binding protein, B9L, in tumorigenesis induced by aberrant activation of Wnt signaling. *Cancer Res* 64: 8496–8501.
- Andersen CL, Christensen LL, Thorsen K, Schepeler T, Sørensen FB, Verspaget HW, Simon R, Kruhøffer M, Aaltonen LA, Laurberg S, et al. 2009. Dysregulation of the transcription factors SOX4, CBFB and SMARCC1 correlates with outcome of colorectal cancer. *Br J Cancer* 100: 511–523.
- Batthey J, Moulding C, Taub R, Murphy W, Stewart T, Potter H, Lenoir G, Leder P. 1983. The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34: 779–787.
- Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98: 387–396.
- Buttyn R, Sawczuk IS, Benson MC, Siegal JD, Olsson CA. 1987. Enhanced expression of the c-myc protooncogene in high-grade human prostate cancers. *Prostate* 11: 327–337.
- Cher ML, Bova GS, Moore DH, Small EJ, Carroll PR, Pin SS, Epstein JI, Isaacs WB, Jensen RH. 1996. Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res* 56: 3091–3102.
- Chung S, Nakagawa H, Uemura M, Piao L, Ashikawa K, Hosono N, Takata R, Akamatsu S, Kawaguchi T, Morizono T, et al. 2011. Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Sci* 102: 245–252.
- Couzin J, Kaiser J. 2007. Genome-wide association. Closing the net on common disease genes. *Science* 316: 820–822.
- De la Roche M, Worm J, Bienz M. 2008. The function of BCL9 in Wnt/beta-catenin signaling and colorectal cancer cells. *BMC Cancer* 8: 199.
- De Laat W, Dekker J. 2012. 3C-based technologies to study the shape of the genome. *Methods* 58: 189–191.
- De Laat W, Van De Werken HJG, De Vree PJP, Splinter E, Holwerda SJB, Klous P, De Wit E. 2012. *4C technology: Protocols and data analysis*. 1st ed. Elsevier Inc.
- Dekker J, Phillips-cremins JE, Sauria MEG, Sanyal A, Tatiana I, Lajoie BR, Bell JSK, Ong C, Hookway TA, Guo C, et al. 2014. Architectural protein subclasses shape 3-D organization of genomes during lineage commitment. *Cell* 153: 1281–1295.
- Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295: 1306–1311.
- DelBove J, Rosson G, Strobeck M, Chen J, Archer TK, Wang W, Knudsen ES, Weissman BE. 2011. Identification of a core member of the SWI/SNF complex, BAF155/SMARCC1, as a human tumor suppressor gene. *Epigenetics* 6: 1444–1453.
- Deng K, Guo X, Wang H, Xia J. 2014. The lncRNA-MYC regulatory network in cancer. *Tumour Biol* 35: 9497–9503.
- Dorsett D. 2011. Cohesin: genomic insights into controlling gene transcription and development. *Curr Opin Genet Dev* 21: 199–206.
- Douglas EJ, Fiegler H, Rowan A, Halford S, Bicknell DC, Bodmer W, Tomlinson IPM, Carter NP. 2004. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 64: 4817–4825.
- Easton DF, Eeles R a. 2008. Genome-wide association studies in cancer. *Hum Mol Genet* 17: R109–115.
- Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093.
- Evan G, Daniel JM, Junttila MR, Pouyet L, Karnezis A, Bui DA, Brown-swigart L, Johnson L, Gerard I. 2009. Distinct thresholds govern Myc's biological output in vivo. *Cancer Cell* 14: 447–457.

- Fan J, Liu J, Culty M, Papadopoulos V. 2010. Acyl-coenzyme A binding domain containing 3 (ACBD3; PAP7; GCP60): an emerging signaling molecule. *Prog Lipid Res* 49: 218–234.
- Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, Xu H, Ramsay RG, Odom DT, Flicek P. 2012. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis -regulatory modules. *Genome Res* 22: 2163–2175.
- Finocchiaro G, Mancuso F, Muller H. 2005. Mining published lists of cancer related microarray experiments: identification of a gene expression signature having a critical role in cell-cycle control. *BMC Bioinformatics* 6 Suppl 4: S14.
- Fraser P, Bickmore W. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature* 447: 413–417.
- Freedman ML, Haiman C a, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, et al. 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 103: 14068–14073.
- Frietze S, Wang R, Yao L, Tak Y, Ye Z, Gaddis M, Witt H, Farnham PJ, Jin VX. 2012. Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol* 13: R52.
- Fullwood MJ, Wei C-L, Liu ET, Ruan Y. 2009. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* 19: 521–532.
- Gil J, Kerai P, Leonart M, Bernard D, Cigudosa JC, Peters G, Carnero A, Beach D. 2005. immortalization of primary human prostate epithelial cells by c-Myc. *Cancer Res* 65: 2179–2185.
- Göndör A, Rougier C, Ohlsson R. 2008. High-resolution circular chromosome conformation capture assay. *Nat Protoc* 3: 303–313.
- Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, et al. 2007. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39: 631–637.
- Gurel B, Iwata T, Koh CM, Jenkins RB, Lan F, Van Dang C, Hicks JL, Morgan J, Cornish TC, Sutcliffe S, et al. 2008. Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. *Mod Pathol* 21: 1156–1167.
- Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forné T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2: 1722–1733.
- Haiman C a, Le Marchand L, Yamamoto J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE. 2007. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 39: 954–956.
- Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E, Taipale J. 2006. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* 124: 47–59.
- Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CWH, Ye C, Ping JLH, Mulawadi F, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* 43: 630–638.
- Hazelett DJ, Rhie SK, Gaddis M, Yan C, Lakeland DL, Coetzee SG, Henderson BE, Noushmehr H, Cozen W, Kote-Jarai Z, et al. 2014. Comprehensive Functional Annotation of 77 Prostate Cancer Risk Loci. *PLoS Genet* 10: e1004102.
- He TC, Sparks a B, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW. 1998. Identification of c-MYC as a target of the APC pathway. *Science* 281: 1509–1512.
- Heebøll S, Borre M, Ottosen PD, Andersen CL, Mansilla F, Dyrskjøt L, Orntoft TF, Tørring N. 2008. SMARCC1 expression is upregulated in prostate cancer and positively correlated with tumour recurrence and dedifferentiation. *Histol Histopathol* 23: 1069–1076.
- Hong J-W, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* 321: 1314.
- Horoszewicz JS, Leong SS, Kawinski E, Karr JP, Rosenthal H, Chu TM, Mirand E a, Murphy GP. 1983. LNCaP Model of Human Prostatic Carcinoma LNCaP Model of Human Prostatic Carcinoma1. *Cancer Res* 43: 1809–1818.

- Huppi K, Volfovsky N, Runfola T, Jones TL, Mackiewicz M, Martin SE, Mushinski JF, Stephens R, Caplen NJ. 2008. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol Cancer Res* 6: 212–221.
- Jackson V. 1999. Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods* 17: 125–139.
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov V V, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128: 1231–1245.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra M a. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res* 19: 1639–1645.
- Lapointe J, Li C, Giacomini CP, Salari K, Huang S, Wang P, Ferrari M, Hernandez-Boussard T, Brooks JD, Pollack JR. 2007. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res* 67: 8504–8510.
- Lavenu A, Pournin S, Babinet C, Morello D. 1994. The cis-acting elements known to regulate c-myc expression ex vivo are not sufficient for correct transcription in vivo. *Oncogene* 9: 527–536.
- Lerner I, Baraz L, Pikarsky E, Meirovitz A, Edovitsky E, Peretz T, Vlodavsky I, Elkin M. 2008. Function of heparanase in prostate tumorigenesis: potential for therapy. *Clin Cancer Res* 14: 668–676.
- Levens D. 2010. “You Don’t Muck with MYC.” *Genes Cancer* 1: 547–554.
- Lomvardas S, Barnea G, Pisapia D, Mendelsohn M, Kirkland J, Axel R. 2006. Interchromosomal Interactions and Olfactory Receptor Choice. *Cell* 126: 403–413.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Ling H, Spizzo R, Atlasi Y, Nicoloso M, Shimizu M, Redis RS, Nishida N, Gafà R, Song J, Guo Z, et al. 2013. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res* 23: 1446–1461.
- Lovén J, Hoke H a., Lin CY, Lau A, Orlando D a., Vakoc CR, Bradner JE, Lee TI, Young R a. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153: 320–334
- Lowe SW, Cepero E, Evan G. 2004. Intrinsic tumour suppression. *Nature* 432: 307–315.
- Mani M, Carrasco DE, Yunyu Z, Takada K, Gatt ME, Dutta-Simmons J, Ikeda H, Diaz-Griffero F, Pena-Cruz V, Bertagnolli M, et al. 2009. BCL9 promotes tumor progression by conferring enhanced proliferative, metastatic, and angiogenic properties to cancer cells. *Cancer Res* 69: 7577–7586.
- Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* 128: 787–800.
- Müller WG, Rieder D, Karpova TS, John S, Trajanoski Z, McNally JG. 2007. Organization of chromatin and histone modifications at a transcription site. *J Cell Biol* 177: 957–967.
- Naumova N, Smith EM, Zhan Y, Dekker J. 2012. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* 58: 192–203.
- Nissan A, Stojadinovic A, Mitrani-Rosenbaum S, Halle D, Grinbaum R, Roistacher M, Bochem A, Dayanc BE, Ritter G, Gomceli I, et al. 2012. Colon cancer associated transcript-1: a novel RNA expressed in malignant and pre-malignant human tissues. *Int J Cancer* 130: 1598–1606.
- Ong C-T, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 15: 234–246.
- Orlando V, Strutt H, Paro R. 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* 11: 205–214.
- Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36: 1065–1071.
- Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132: 422–433.
- Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. *Cell* 137: 1194–1211.

- Pombo A, Jackson DA, Hollinshead M, Wang Z, Roeder RG, Cook PR. 1999. Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J* 18: 2241–2253.
- Pomerantz M, Jia L, Landan G, Jaschek R, Herman P, Reich D, Yan C, Khalid O, Kantoff P, Oh W, et al. 2009. Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet* 5: e1000597.
- Pomerantz MM, Beckwith CA, Regan MM, Wyman SK, Chen Y, Hawksworth DJ, Schumacher FR, Penney KL, Stampfer MJ, Chan JA, et al. 2010. expression. 69: 5568–5574.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, et al. 2007. OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9: 166–180.
- Ribeiro FR, Jerónimo C, Henrique R, Fonseca D, Oliveira J, Lothe RA, Teixeira MR. 2006. 8q gain is an independent predictor of poor survival in diagnostic needle biopsies from prostate cancer suspects. *Clin Cancer Res* 12: 3961–3970.
- Ried T, Knutzen R, Steinbeck R, Blegen H, Schröck E, Heselmeyer K, du Manoir S, Auer G. 1996. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 15: 234–245.
- Rubio ED, Reiss DJ, Welcsh PL, Distèche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105: 8309–8314.
- Sakamoto I, Ohwada S, Toya H, Togo N, Kashiwabara K, Oyama T, Nakajima T, Ito H, Adachi S, Jigami T, et al. 2007. Up-regulation of a BCL9-related beta-catenin-binding protein, B9L, in different stages of sporadic colorectal adenoma. *Cancer Sci* 98: 83–87.
- Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* 489: 109–113.
- Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* 20: 578–588.
- Schriml LM, Padilla-Nash HM, Coleman a., Moen P, Nash WG, Menninger J, Jones G, Ried T, Dean M. 1999. Tyramide signal amplification (TSA)-FISH applied to mapping PCR-labeled probes less than 1 kb in size. *Biotechniques* 27: 608–613.
- Shadeo A, Chari R, Lonergan KM, Pusic A, Miller D, Ehlen T, Van Niekerk D, Maticic J, Richards-Kortum R, Follen M, et al. 2008. Up regulation in gene expression of chromatin remodelling factors in cervical intraepithelial neoplasia. *BMC Genomics* 9: 64.
- Shen MM, Abate-Shen C. 2010. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes and Dev*: 24: 1967–2000.
- Shin H, Liu T, Manrai AK, Liu SX. 2009. CEAS: Cis-regulatory element annotation system. *Bioinformatics* 25: 2605–2606.
- Simonis M, Kooren J, de Laat W. 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4: 895–901.
- Smith TF, Waterman MS, Fitch WM. 1981. Comparative biosequence metrics. *J Mol Evol* 18: 38–46.
- Solomon MJ, Varshavsky A. 1985. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A* 82: 6470–6474.
- Song Z, Xu X, Wei S, Chen J, Zhou N. 2008. [The expression and clinical significance of heparanase in non-small cell lung cancer.]. *Zhongguo Fei Ai Za Zhi* 11: 798–801.
- Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Stephens RM, Harris TJR, Munroe DJ, Wu X. 2010. Long-range enhancers on 8q24 regulate c-Myc. *Proc Natl Acad Sci U S A* 107: 3001–3005.
- Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell R a. 2005. Interchromosomal associations between alternatively expressed loci. *Nature* 435: 637–645.
- Sun J, Liu W, Adams TS, Sun J, Li X, Turner AR, Chang B, Kim JW, Zheng SL, Isaacs WB, et al. 2007. DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. *Prostate* 67: 692–700.

- Sun T, Co NN, Wong N. 2014. PFTK1 interacts with cyclin Y to activate non-canonical Wnt signaling in hepatocellular carcinoma. *Biochem Biophys Res Commun* 449: 163–168.
- Takada K, Zhu D, Bird GH, Sukhdeo K, Zhao J-J, Mani M, Lemieux M, Carrasco DE, Ryan J, Horst D, et al. 2012. Targeted disruption of the BCL9/ $\beta$ -catenin complex inhibits oncogenic Wnt signaling. *Sci Transl Med* 4: 148ra117.
- Tan K, He B, Chen C, Teng L. 2014. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* 111: E2191–E2199.
- Tang D, Zhang Q, Zhao S, Wang J, Lu K, Song Y, Zhao L, Kang X, Wang J, Xu S, et al. 2013. The expression and clinical significance of microRNA-1258 and heparanase in human breast cancer. *Clin Biochem* 46: 926–932.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178–192.
- Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, et al. 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39: 984–988.
- Toujani S, Dessen P, Ithzar N, Danglot G, Richon C, Vassetzky Y, Robert T, Lazar V, Bosq J, Da Costa L, et al. 2009. High resolution genome-wide analysis of chromosomal alterations in Burkitt's lymphoma. *PLoS One* 4: e7089.
- Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Björklund M, Wei G, Yan J, Niittymäki I, et al. 2009. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41: 885–890.
- Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner, Griffin C, Kallioniemi O, Visakorpi, McGill, Herath J. 1995. DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications. *Clin Cancer Res* 1: 11–18.
- Van Steensel B, Dekker J. 2010. Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* 28: 1089–1095.
- Visakorpi T, Kallioniemi AH, Syvänen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP. 1995. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res* 55: 342–347.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457: 854–858.
- Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474: 390–394.
- Wang L, Du M, Yuan T, Schilter KF, Dittmar RL, Mackinnon A, Tschannen M, Worthey E, Jacob H, Xia S, et al. 2014a. Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome. *Hum Mol Gen* 24: 154–166.
- Wang L, Liu R, Li W, Chen C, Katoh H, Chen G-Y, McNally B, Lin L, Zhou P, Zuo T, et al. 2009. Somatic single hits inactivate the X-linked tumor suppressor FOXP3 in the prostate. *Cancer Cell* 16: 336–346.
- Wang L, Zhao Z, Meyer MB, Saha S, Yu M, Guo A, Wisinski KB, Huang W, Cai W, Pike JW, et al. 2014b. CARM1 methylates chromatin remodeling factor BAF155 to enhance tumor progression and metastasis. *Cancer Cell* 25: 21–36.
- Warner MH, Roinick KL, Arndt KM. 2007. Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. *Mol Cell Biol* 27: 6103–6115.
- Wasserman NF, Aneas I, Nobrega M a. 2010. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res* 20: 1191–1197.
- Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451: 796–801.
- West AG, Fraser P. 2005. Remote control of gene transcription. *Hum Mol Genet* 14 Spec No: R101–111.
- Wierstra I, Alves J. 2008. The c-myc promoter: still MysterY and challenge. *Adv Cancer Res* 99: 113–333.

- Williams K, Fernandez S, Stien X, Ishii K, Love HD, Lau Y-FC, Roberts RL, Hayward SW. 2005. Unopposed c-MYC expression in benign prostatic epithelium causes a cancer phenotype. *Prostate* 63: 369–384.
- Willis TG, Zalcborg IR, Coignet LJ, Wlodarska I, Stul M, Jadayel DM, Bastard C, Treleaven JG, Catovsky D, Silva ML, et al. 1998. Molecular cloning of translocation t(1;14)(q21;q32) defines a novel gene (BCL9) at chromosome 1q21. *Blood* 91: 1873–1881.
- Witte JS. 2007. Multiple prostate cancer risk variants on 8q24. *Nat Genet* 39: 579–580.
- Xiang J-F, Yin Q-F, Chen T, Zhang Y, Zhang X-O, Wu Z, Zhang S, Wang H-B, Ge J, Lu X, et al. 2014. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24: 513–531.
- Xu M, Cook PR. 2008. Similar active genes cluster in specialized transcription factories. *J Cell Biol* 181: 615–623.
- Yang F, Xue X, Bi J, Zheng L, Zhi K, Gu Y, Fang G. 2013. Long noncoding RNA CCAT1, which could be activated by c-Myc, promotes the progression of gastric carcinoma. *J Cancer Res Clin Oncol* 139: 437–445.
- Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, et al. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39: 645–649.
- Yochum GS, Cleland R, Goodman RH. 2008. A genome-wide screen for beta-catenin binding sites identifies a downstream enhancer element that controls c-Myc gene expression. *Mol Cell Biol* 28: 7368–7379.
- Yochum GS, McWeeney S, Rajaraman V, Cleland R, Peters S, Goodman RH. 2007. Serial analysis of chromatin occupancy identifies beta-catenin target genes in colorectal carcinoma cells. *Proc Natl Acad Sci U S A* 104: 3324–3329.
- Yu J, Yu J, Mani R-S, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, et al. 2010. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 17: 443–454.
- Yu X, Wang Y, Jiang M, Brier B, Roy-Burman P, Shen MM, Taketo MM, Wills M, Matusik RJ. 2009. Activation of beta-Catenin in mouse prostate causes HGPIN and continuous prostate growth after castration. *Prostate* 69: 249–262.
- Yue W, Zhao X, Zhang L, Xu S, Liu Z, Ma L, Jia W, Qian Z, Zhang C, Wang Y, et al. 2011. Cell cycle protein cyclin Y is associated with human non-small-cell lung cancer proliferation and tumorigenesis. *Clin Lung Cancer* 12: 43–50.
- Zanke BW, Greenwood CMT, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, et al. 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. 39: 989–994.
- Zhang W, Chan H, Wei L, Pan Z, Zhang J, Li L. 2013. Overexpression of heparanase in ovarian cancer and its clinical significance. *Oncol Rep* 30: 2279–2287.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Open Access Model-based Analysis of CHIP-Seq (MACS). *Gen Biol* 9: R137.
- Zhao Z, Go A, Mariano P, Wang S, Tavoosidana G, Sjo M, Kanduri C, Lezcano M, Sandhu KS, Singh U, et al. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. 38: 1341–1348.

## LISTA DE FIGURAS

Figura 1 – Mapa físico de la región 8q24.....	3
Figura 2 – Interacciones de largo alcance.....	7
Figura 3 – Esquema de la metodología 3C.....	9
Figura 4 – Esquema de la metodología 4C.....	10
Figura 5 – Metodología 4C en detalle.....	15
Figura 6 – Secuenciación en Ion Torrent.....	22
Figura 7 – Procesamiento de los datos 4C.....	23
Figura 8 – Interpretación funcional de picos 4C.....	27
Figura 9 – Cebadores para determinar la eficiencia de la digestión .....	31
Figura 10 – Cebadores para el enriquecimiento 4C.....	32
Figura 11 – Cebadores para validación por 3C-PCR.....	33
Figura 12 – Eficiencia de la digestión de la cromatina.....	36
Figura 13 – Enriquecimiento de bibliotecas 4C por PCR invertido y anidado.....	37
Figura 14 – Fragmentación de los amplicones 4C del potenciador E.....	38
Figura 15 – Perfiles de las librerías de secuenciación de las bibliotecas 4C enriquecidas.....	39
Figura 16 – Remoción de dímeros de adaptadores de una librería de secuenciación.....	40
Figura 17 – Primeros resultados de la secuenciación.....	41
Figura 18 – Distribución del largo de las lecturas de las bibliotecas 4C .....	42
Figura 19 – Homología en el genoma humano de lecturas quiméricas a través de BLAST.....	44
Figura 20 – Homología en el genoma humano de lecturas quiméricas a través de BLAST.....	45
Figura 21 – Distribución de picos 4C en el genoma.....	49
Figura 22 – Distribución de los picos 4C en características genómicas.....	50
Figura 23 – Promedio de picos 4C respecto a un metagen.....	51
Figura 24 – Distribución de picos 4C en estructuras génicas.....	52
Figura 25 – Intersección de picos 4C del potenciador E con marcadores de promotores.....	53
Figura 26 – Intersección de picos 4C de c-Myc con marcadores de potenciadores.....	57
Figura 27 – Colocalización de un pico 4C de c-Myc y marcadores funcionales.....	59
Figura 28 – Colocalización de un pico 4C del potenciador E y marcadores funcionales.....	60
Figura 29 – Contactos genómicos del potenciador E.....	61
Figura 30 – Contactos genómicos del promotor de c-Myc.....	62
Figura 31 – Confirmación de contactos 4C del potenciador E a través de 3C-PCR.....	63
Figura 32 – Lecturas quiméricas 4C que contienen al gen ACBD3 y al potenciador E.....	64
Figura 33 – Lecturas quiméricas 4C que contienen al gen SMARCC1 y al potenciador E.....	65
Figura 34 – Confirmación de contactos 4C de c-Myc a través de 3C-PCR.....	65
Figura 35 – Lecturas quiméricas 4C conteniendo al potenciador candidato 9p23 y a c-Myc.....	66
Figura 36 – Comparación de las interacciones de 8q24 en LNCaP en otro estudio de captura...67	

Figura 37 – Distribución genómica de elementos potenciadores en otras líneas celulares.....	73
Figura 38 – Marcadores de elementos potenciadores en picos 4C de c-Myc.....	75
Figura 39 – Potenciadores que interactúan con uno o más promotores.....	80
Figura 40 – Promotores que interactúan con uno o más potenciadores.....	80

## LISTA DE TABLAS

Tabla 1 – Datos de secuenciación utilizados para intersectar con datos 4C.....	26
Tabla 2 – Cebadores utilizados para medir la eficiencia de la digestión.....	31
Tabla 3 – Cebadores utilizados para el enriquecimiento durante 4C.....	32
Tabla 4 – Cebadores utilizados para 3C-PCR.....	36
Tabla 5 – Secuenciación de las bibliotecas 4C.....	41
Tabla 6 – Procesamiento y alineamiento de las lecturas.....	42
Tabla 7 – Resultados del alineamiento.....	43
Tabla 8 – Lecturas quiméricas con distintas fracciones de secuencias .....	43
Tabla 9 – Ontología de los genes más cercanos a los picos 4C en común a ambas carnadas. .	46
Tabla 10 – Picos 4C con mayor número de lecturas para cada carnada.....	47
Tabla 11 – Ontología de los genes más cercanos a los picos 4C más abundantes.....	48
Tabla 12 – Genes con picos 4C del potenciador E y H3K4me3 en su región promotora.....	54
Tabla 13 – Marcadores funcionales en picos 4C del potenciador E (H3K4me3) .....	55
Tabla 14 – Genes seleccionadas como candidatos regulados por el potenciador E.....	55
Tabla 15 – Marcadores funcionales en picos 4C de c-Myc (H3K4me1 y H3K27ac).....	57
Tabla 16 – Marcadores funcionales en picos 4C de c-Myc (p300).....	58
Tabla 17 – Regiones seleccionadas como potenciadores candidatos a interactuar con c-Myc. .	58
Tabla 18 – Comparación de las interacciones de 8q24 con otro estudio de captura.....	68
Tabla 19 – Marcadores de cromatina abierta en genes y potenciadores candidatos.....	76
Tabla 20 – Marcadores de lazos de cromatina en genes y potenciadores candidatos.....	77