



Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Biológicas
Programa de Doctorado en Ciencias Biológicas
Área: Biología Celular y Molecular

Arquitectura regulatoria osteogénica del locus SCPP



Tesis para obtener el grado de Doctor en Ciencias Biológicas
Área: Biología Celular y Molecular

Adrián Romero Mora
Concepción-Chile
2021

Profesor guía: Dr. Sylvain Marcellini Liotaud
Dpto. de Biología Celular
Facultad de Ciencias Biológicas
Universidad de Concepción

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

Profesores integrantes Comisión Evaluadora:

Dr. Sylvain Marcellini Liotaud
Profesor Guía de Tesis
Facultad de Ciencias Biológicas
Universidad de Concepción

Dr. José L. Gutiérrez C.
Dpto. Bioquímica y Biología Molecular
Facultad de Ciencias Biológicas
Universidad de Concepción



Dra. María Inés Barría C.
Dpto. Microbiología
Facultad de Ciencias Biológicas
Universidad de Concepción

Dr. Ariel Reyes Z.
Profesor Evaluador Externo
Universidad Andrés Bello

Dr. Ariel Castro Alma
Director
Programa Doctorado en Ciencias Biológicas
Área Biología Celular y Molecular

Tesis de doctorado financiada por:

FONDECYT regular # 1151196

FONDECYT regular # 1190926

ECOS-Sud C15S02

Beca ANID doctorado nacional # 21160069

Beca ANID gastos operacionales 2019

Beca ANID extensión de beca para redacción de tesis doctoral 2020

Beca décimo semestre proyecto EDPG



DEDICATORIA

A mi padre, Don Víctor Romero y a mi recién llegada hija María Emilia, quienes no se conocieron en vida, pero probablemente se encontraron en el camino de la inmensidad del cosmos.

A mis padres, por la vida, por haberme apoyado en mi decisión de estudiar una carrera en ciencias, por su fuerza y sus enseñanzas; y por tolerar los años que me he ausentado de casa.

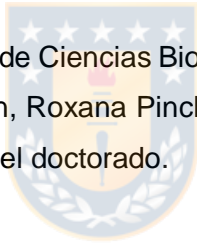
A mi amada esposa Dávila Vitoriano, quien llegó a mi vida justo cuando emprendía este proyecto y me ha ayudado a cumplir mis sueños. Su apoyo en lo académico, moral y espiritual ha sido imprescindible.

AGRADECIMIENTOS

Al mi tutor de tesis, el Dr. Sylvain Marcellini, por haberme aceptado en su laboratorio a pesar de mi inactividad en la ciencia durante varios años. Gracias por confiar en mí y por ayudarme a retomar mi carrera científica; por ayudarme a escribir postulaciones a becas y apoyos para congresos; también por sus intensas enseñanzas en el mundo de la regulación transcripcional y la evolución. Expreso también mi gratitud por su amistad, la cual considero sumamente valiosa.

A mis colegas del LADE. Los de antes (Ariana, Patricia, Daniel y Javier), porque sus trabajos fueron fundamentales para iniciar con este proyecto. A los colegas más recientes, particularmente a David Muñoz, por enseñarme las técnicas de biología molecular indispensables en el laboratorio; a Héctor Castillo, por enseñarme a procesar los datos extraídos del citómetro de flujo y por sus datos de ATAC-seq. A Francisco, por su apoyo puntual en temas de bioinformática y a Fret por recibirme cuando llegué a Chile y por brindarme su amistad.

Al claustro de profesores de la Facultad de Ciencias Biológicas, particularmente a los profesores Juan Pablo Henríquez, Marcela Torrejón, Roxana Pinchera y Ariel Castro. Por sus clases y por su valioso apoyo académico a lo largo del doctorado.



A los profesores que integran la comisión de tesis, por sus importantes comentarios y sugerencias cuando realicé la defensa del proyecto y también por tomarse el tiempo para revisar y enriquecer este trabajo.

A Margareth Santana, por su imprescindible apoyo en los procesos administrativos y por su amistad.

A la gente de Francia, principalmente a Charlotte Andrieu-Soler, quien me ayudó a diseñar el 4C y al Dr. Eric Soler, quien me permitió hacer una estancia en su laboratorio y gestionó para que Mohammad Salma me ayudara a hacer los análisis bioinformáticos. También agradezco a Daan Noordermeer quien, sin conocerme, me ayudó a resolver dificultades técnicas que enfrenté en la estandarización del protocolo de 4C.

Al personal que integra el CMA, principalmente a German y Asia, por ayudarme a obtener las imágenes de HIS y por su valiosa amistad.

Al pueblo chileno, de quienes se obtuvieron los recursos para que el gobierno del país me asignara la beca ANID. Sin ello, hubiera sido imposible mi estadía en Chile. De algunos chilenos recibí apoyo directo. Principalmente Andrés y José, quienes me ayudaron a hacer súper cómodos los años 2016-2017.

A mis amigos Claudia Álvarez, Sol Torruella, Cristian Gidi y Eder Ramírez por brindarme su amistad. Particularmente expreso mi gratitud para mi amigo Patricio Ordenes, quien me apoyó en todo momento con absoluta incondicionalidad.

A mis profesores de México: Marcelino Cerejido, Arturo Ponce y Hugo Beraldi, quienes confiaron en mí y enviaron las cartas de recomendación para mi proceso de ingreso al doctorado.

A mis amigos mexicanos Luis Osorio, Augusto Alonso y Norberto Hernández, por el apoyo que me brindaron cuando viví en la Ciudad e México y por impulsarme a continuar en la ciencia.

A mis hermanos, Juan, Ismael y Agustín: por apoyarme cuando estuvo en sus manos durante mi largo camino de estudiante. En mi niñez, Ismael me ayudaba a forrar mis libros. Juan me inculcó el gusto por la lectura y me ayudaba con las matemáticas y Agustín, era mi figura de super héroe. Aunque estuvieron relativamente ausentes durante mi formación doctoral, no dejan de ser parte fundamental en mi vida.

Finalmente, a todos los que, sin mencionarlos aquí, de alguna forma estuvieron apoyándome en este proceso.

TABLA DE CONTENIDOS

DEDICATORIA	3
AGRADECIMIENTOS	4
ABREVIATURAS	10
GLOSARIO	12
RESUMEN	14
ABSTRACT	16
1. INTRODUCCIÓN	18
1.1 DUPLICACIONES LOCALES Y FAMILIAS DE GENES ORGANIZADOS EN TÁNDEM	18
1.2 REGULACIÓN TRANSCRIPCIONAL.....	20
1.3 REGULACIÓN EPIGENÉTICA DE LA EXPRESIÓN GÉNICA.....	22
1.4 DOMINIOS DE ASOCIACIÓN TOPOLÓGICA (TADs).....	23
1.5 ARQUITECTURA REGULATORIA EN VERTEBRADOS.....	27
1.6 PATRONES DE EXPRESIÓN Y FUNCIÓN DE LOS MIEMBROS SCPP Y SPARC.....	32
1.6.1 <i>DMP1</i>	34
1.6.2 <i>BSP</i>	35
1.6.3 <i>OPN</i>	35
1.6.4 <i>SCPPA2</i>	36
1.6.5 <i>SPARC</i>	36
1.7 REGULACIÓN TRANSCRIPCIONAL DE LOS MIEMBROS SCPP.....	37
1.8 ARQUITECTURA REGULATORIA EN OSTEÓBLASTOS.....	39
2. PREGUNTA DE INVESTIGACIÓN E HIPÓTESIS	42
3. OBJETIVO GENERAL	44
4. OBJETIVOS ESPECÍFICOS	44
5. MATERIALES Y MÉTODOS	45
5.1 ANIMALES DE EXPERIMENTACIÓN Y PREPARACIÓN DE MUESTRAS	45
5.2 ANÁLISIS DE EXPRESIÓN ESPACIAL POR HIBRIDACIÓN <i>IN SITU</i>	45
5.2.1 <i>Amplificación de regiones intrónicas</i>	45
5.2.2 <i>Reacciones de transcripción in vitro</i>	46
5.2.3 <i>Hibridación in situ con sondas intrónicas y exónicas</i>	46
5.3 SELECCIÓN DE MCRS PUTATIVOS	47
5.4 GENERACIÓN DE PLÁSMIDOS PARA ENSAYOS REPORTEROS	48
5.4.1 <i>Amplificación y purificación de promotores basales y enhancers</i>	48
5.4.2 <i>Ligación, transformación y purificación de plásmidos CisTA</i>	48
5.4.3 <i>Generación de vectores ENTRY</i>	49
5.4.4 <i>Reacción BP</i>	50
5.4.5 <i>Ligación de enhancers en el vector Entry</i>	50
5.4.6 <i>Reacción LR</i>	51
5.5 ENSAYOS FUNCIONALES CON REPORTEROS.....	53
5.5.1 <i>Estandarización de electroporación mediante el sistema Neon</i>	53
5.5.2 <i>Cultivo primario de osteoblastos</i>	56
5.5.3 <i>Electroporación de reporteros del locus SCPP y SPARC</i>	56
5.5.4 <i>Cuantificación de fluorescencia por citometría de flujo</i>	57
5.5.5 <i>Análisis de datos de citometría</i>	59

5.5.6	<i>Análisis estadístico</i>	59
5.6	ENSAYO DE CAPTURA DE CONFORMACIÓN DE CROMOSOMA CIRCULAR (4C-SEQ)	61
5.6.1	<i>Selección de enzimas</i>	61
5.6.2	<i>Diseño de partidores</i>	64
5.6.3	<i>Amplificación y ligación de Viewpoints</i>	66
5.6.4	<i>Preparación de la muestra</i>	67
5.6.5	<i>Crosslinking</i>	67
5.6.6	<i>Lisis celular</i>	67
5.6.7	<i>Primera digestión</i>	70
5.6.8	<i>Primera ligación y purificación del templado (3C)</i>	70
5.6.9	<i>Segunda digestión</i>	71
5.6.10	<i>Segunda ligación y purificación del templado (4C)</i>	72
5.6.11	<i>Generación de librerías (4C) por PCR inversa</i>	73
5.6.12	<i>Procesamiento de datos</i>	76
6.	RESULTADOS	80
6.1	EXPRESIÓN ESPACIAL ESPECÍFICA DE LOS MIEMBROS DEL LOCUS SCPP Y SPARC	80
6.1.1	<i>DMP1 se expresa específicamente en osteocitos</i>	83
6.1.2	<i>IBSP se expresa específicamente en osteoblastos</i>	85
6.1.3	<i>Sparc se expresa específicamente en osteoblastos</i>	87
6.1.4	<i>SCPPA2 tiene un patrón de expresión divergente</i>	90
6.2	ANÁLISIS FUNCIONAL DE LOS MCRs DEL LOCUS SCPP Y DE SPARC.....	92
6.2.1	<i>Expresión de miembros del locus SCPP y SPARC en hueso</i>	92
6.2.2	<i>Paisaje regulatorio del locus SCPP</i>	94
6.2.3	<i>MCRs putativos ligados al gen Sparc</i>	97
6.2.4	<i>Un enhancer proximal a IBSP activa su expresión en osteoblastos</i>	99
6.2.5	<i>Dos MCRs proximales al promotor de Scppa2 muestran actividad silenciadora</i>	101
6.2.6	<i>Los enhancers putativos del locus SCPP no alteran la expresión de Dmp1</i>	103
6.2.7	<i>Dos MCRs proximales al promotor de Sparc activan la transcripción</i>	104
6.3	PERFIL DE INTERACCIONES DE LOS MIEMBROS SCPP Y SPARC.....	107
6.3.1	<i>Paisaje de contactos del gen lbsp</i>	108
6.3.2	<i>lbsp contacta al promotor del Dmp1</i>	112
6.3.3	<i>Paisaje de interacciones del gen Scppa2</i>	115
6.3.4	<i>Paisaje de interacciones del gen Sparc</i>	120
6.3.5	<i>Todos contra todos</i>	124
7.	DISCUSIÓN	126
7.1	EXPRESIÓN DE LOS GENES SCPP Y SPARC (UNA ACTUALIZACIÓN).....	127
7.2	ARQUITECTURA REGULATORIA DEL LOCUS SCPP	128
7.3	DOS PROMOTORES QUE INTERACTÚAN SE EXPRESAN DIFERENCIALMENTE	129
7.4	ARQUITECTURA REGULATORIA DE SPARC	131
7.5	CTCF EN LA ARQUITECTURA REGULATORIA DEL LOCUS SCPP	134
7.6	EVOLUCIÓN DEL LOCUS SCPP Y SPARC.....	136
8.	CONCLUSIONES	140
9.	BIBLIOGRAFÍA	141
	ANEXO 1: PARTIDORES	151
	TABLA 1: SECUENCIAS DE PARTIDORES PARA DISEÑO DE SONDAS INTRÓNICAS	151
	TABLA 2: SECUENCIAS DE PARTIDORES PARA AMPLIFICAR BAITS	151
	TABLA 3: SECUENCIAS DE PARTIDORES PARA PCR INVERSA	152
	TABLA 4: SECUENCIAS DE PARTIDORES PARA PCR INVERSA CON ADAPTADORES	152

ANEXO 2: INTERACCIONES CON OTROS CROMOSOMAS	153
FIGURA 1: REGIONES DE INTERACCIÓN EN <i>TRANS</i> CON EL PROMOTOR <i>AMBN</i>	153
FIGURA 2: REGIONES DE INTERACCIÓN EN <i>TRANS</i> CON EL PROMOTOR <i>IBSP</i>	153
FIGURA 3: REGIONES DE INTERACCIÓN EN <i>TRANS</i> CON EL PROMOTOR <i>DMP1</i>	154
FIGURA 4: REGIONES DE INTERACCIÓN EN <i>TRANS</i> CON EL PROMOTOR <i>SCPPA2</i>	154
FIGURA 5: REGIONES DE INTERACCIÓN EN <i>TRANS</i> CON EL PROMOTOR <i>SPARC</i>	155
ANEXO 4: SPARC Y TADS.....	157

ÍNDICE DE FIGURAS

FIGURA 1: ORIGEN DEL LOCUS <i>SCPP</i> POR EVENTOS DE DUPLICACIÓN GÉNICA	20
FIGURA 2: MECANISMO DE FORMACIÓN DE TADS	25
FIGURA 3: EJEMPLOS DE ARQUITECTURA REGULATORIA	31
FIGURA 4: ORGANIZACIÓN GENÓMICA DEL LOCUS <i>SCPP</i>	34
FIGURA 5: FUNDAMENTO TEÓRICO DEL ENSAYO <i>4C-SEQ</i>	41
FIGURA 6: GENERACIÓN DE VECTORES DE EXPRESIÓN MEDIANTE EL SISTEMA GATEWAY.....	53
FIGURA 7: CALIBRACIÓN DE LAS CONDICIONES DE ELECTROPORACIÓN	55
FIGURA 8: COMPENSACIÓN DE LA ELECTROPORACIÓN	58
FIGURA 9: ESTANDARIZACIÓN DE LA PRIMERA DIGESTIÓN	63
FIGURA 10: EJEMPLO DEL DISEÑO DE PARTIDORES PARA <i>4C</i>	65
FIGURA 11: AMPLIFICACIÓN DE <i>VIEWPOINTS</i>	66
FIGURA 12: OPTIMIZACIÓN DE LA LISIS CELULAR	69
FIGURA 13: OPTIMIZACIÓN DE PCR INVERSA	74
FIGURA 14: OPTIMIZACIÓN DE PCR INVERSA CON ADAPTADORES.....	75
FIGURA 15: PURIFICACIÓN DE LIBRERÍAS POR EL SISTEMA AMPURE	76
FIGURA 16: NORMALIZACIÓN DE LAS SEÑALES DE INTERACCIÓN.....	79
FIGURA 17: ANTECEDENTE EXPERIMENTAL E HISTOLOGÍA DEL HUESO.....	82
FIGURA 18: EXPRESIÓN ESPACIAL DE <i>DMP1</i> CON Sonda EXÓNICA E INTRÓNICA	85
FIGURA 19: EXPRESIÓN ESPACIAL DE <i>IBSP</i> CON Sonda EXÓNICA E INTRÓNICA	86
FIGURA 20: EXPRESIÓN ESPACIAL DE <i>SPARC</i> CON Sonda EXÓNICA E INTRÓNICA	88
FIGURA 21: FUERTE <i>SWITCH</i> REGULATOIRIO EN LA TRANSICIÓN OSTEÓBLASTO-OSTEOCITO	89
FIGURA 22: PATRONES DE EXPRESIÓN DE <i>SCPPA2</i> CON Sonda EXÓNICA E INTRÓNICA	91
FIGURA 23: EXPRESIÓN DE GENES DEL LOCUS <i>SCPP</i> Y <i>SPARC</i>	94
FIGURA 24: PAISAJE REGULATORIO DEL LOCUS <i>SCPP</i>	97
FIGURA 25: PAISAJE REGULATORIO DE <i>SPARC</i>	98
FIGURA 26: ANÁLISIS FUNCIONAL DE MCRs DE <i>IBSP</i>	101
FIGURA 27: ANÁLISIS FUNCIONAL DE MCRs DE <i>SCPPA2</i>	102
FIGURA 28: ANÁLISIS FUNCIONAL DE MCRs DE <i>DMP1</i>	104
FIGURA 29: ANÁLISIS FUNCIONAL DE MCRs DE <i>SPARC</i>	106
FIGURA 30: PERFIL DE INTERACCIONES DE <i>IBSP</i>	109
FIGURA 31: INTERACCIÓN PROXIMAL AL PROMOTOR DE <i>IBSP</i>	111
FIGURA 32: OTRAS REGIONES CANDIDATAS DE INTERACCIÓN CON EL PROMOTOR <i>IBSP</i>	113
FIGURA 33: LOS PERFILES DE INTERACCIÓN SUGIEREN QUE LOS PROMOTORES DE <i>IBSP</i> Y <i>DMP1</i> SE CONTACTAN.....	114
FIGURA 34: PERFILES DE INTERACCIÓN DE <i>SCPPA2</i>	116
FIGURA 35: EL PROMOTOR DE <i>SCPPA2</i> INTERACTÚA CON EL ÚLTIMO EXÓN DE <i>IBSP</i>	118
FIGURA 36: <i>SCPPA2</i> CONTACTA SIGNIFICATIVAMENTE LOS INTRONES DEL GEN BYSTANDER <i>PKD2</i>	119
FIGURA 37: PERFILES DE INTERACCIÓN DE <i>SPARC</i>	121
FIGURA 38: PERFILES DE INTERACCIÓN PROXIMALES AL PROMOTOR DEL GEN <i>SPARC</i>	124

FIGURA 39: TODOS CONTRA TODOS	125
FIGURA 40: MODELO DE INTERACCIÓN <i>IBSP-DMP1</i>	130
FIGURA 41: RESUMEN DE HALLAZGOS EN EL LOCUS SCPP	133
FIGURA 42: ARQUITECTURA REGULATORIA DEL LOCUS SCPP	135
FIGURE 43: SITUACIÓN ANCESTRAL VS SITUACIÓN ACTUAL DE LOS GENES SCPP	137
FIGURA 44: ¿POR QUÉ LOS GENES SCPP NO SE SEPARARON EN LA EVOLUCIÓN?	139
FIGURA 45: TADS EN CÉLULAS HUMANAS ASOCIADOS AL LOCUS SCPP	156
FIGURA 46: TADS EN CÉLULAS HUMANAS ASOCIADOS AL GEN SPARC	157

ÍNDICE DE TABLAS

TABLE 1: OPTIMIZACIÓN DE LA ELECTROPORACIÓN	54
TABLE 2: PCR INVERSA	73
TABLE 3: RESUMEN DE LA ACTIVIDAD FUNCIONAL DE LOS ENHANCERS A,B Y C DE SPARC EN DIFERENTES PASAJES.....	105



ABREVIATURAS

ADN: Ácido Desoxirribonucleico

AMBN: Ameloblastina

AMEL: Amelogenina

ARN: Ácido Ribonucleico

ATAC-seq: Ensayo de cromatina accesible por transposasa acoplado a secuenciación

C-DOM: dominio centromérico

ChIP-seq: Inmunoprecipitación de la cromatina acoplada a secuenciación

CNI: Control Negativo Intrónico

DMP1: Proteína de matriz de dentina 1

DSPP: Sialo-fosfoproteína de dentina

ENAM: Enamelina

ESC: Células troncales embrionarias

FT: Factores de transcripción

GRBs: Bloques de regulación genómica

H3K27ac: Acetilación de la lisina 27 de la histona 3

H3K27me3: Triple metilación de la lisina 27 de la histona 3

H3K4me1: Mono-metilación de la lisina 4 de la histona 3

HES: Tinción hematoxilina eosina

HiC: Captura de conformación de cromosoma (todos contra todos)

IBSP: Sialoproteína de unión a integrina

Kb: Kilobase (mil pares de bases)

Mb: Megabase (1 millón de pares de bases)

MCRs: Módulos cis-reguladores

MEPE: Fosfoglicoproteína de matriz extracelular

mm: micrómetro

NPC: Células neuroprogenitoras

OPN: Proteína osteopontina (codificada por el gen Spp1)

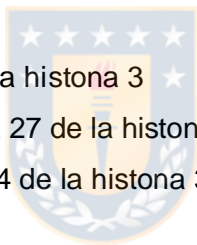
pb: Pares de bases

SCPP: Fosfoproteínas de unión a calcio secretado

SCPPA2: Fosfoproteína ácida de unión a calcio secretado 2

SPARC: Proteína ácida secretada rica en cisteína = Osteonectina (ONC)

T-DOM: dominio telomérico



TADs: Dominios de asociación topológica

VP: Viewpoint

X. tropicalis: *Xenopus tropicalis*

2R: Dos rondas de duplicación del genoma completo

3C: Captura de conformación de cromosoma (1 vs 1)

4C: Captura de conformación de cromosoma circular (1 vs todos)

5C: Captura de conformación de cromosoma con base en la copia de carbono (varios vs varios)



GLOSARIO

Ameloblastos: Células de origen epitelial productoras de esmalte.

Anfioxos: Animales cefalocordados invertebrados. Poseen el plan corporal primitivo de los vertebrados y son usados frecuentemente en evolución de genomas para entender la transición invertebrado-vertebrado. Su genoma evoluciona lentamente y no se ha duplicado ni re-arreglado masivamente como en vertebrados.

Condriccios: Clase de vertebrados con mandíbulas y esqueletos hechos de cartílago en lugar de hueso. Tiburones y rayas son algunos ejemplos.

Condrocitos: Células responsables de la formación del cartílago. El cartílago es fundamental para el desarrollo del hueso en el proceso de osificación endocondral.

Hidroxiapatita: Componente inorgánico del tejido óseo. Consiste predominantemente de calcio y fosfato en un sistema de cristal hexagonal. Fórmula $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$

Lampreas: Familia de agnatos o peces sin mandíbulas con esqueletos cartilaginosos. Se consideran los vertebrados más primitivos.

Odontoblastos: Células que se organizan en la periferia de la pulpa dental. Sintetizan los componentes de la predentina y son responsables de su mineralización. Sintetizan la dentina primaria durante la organogénesis, la dentina secundaria durante la vida del diente y la terciaria cuando el diente sufre estímulos patológicos.

Ortólogos: Genes que fueron derivados por eventos de especiación. En general, se asume que estos genes retienen funciones equivalentes en diferentes organismos y comparten algunas características clave.

Osteíctios: Comúnmente llamados peces óseos. Dentro de este taxón se encuentran los actinopterigios (la clase más grande de vertebrados o teleósteos) y los sarcopterigios (celacantos y peces pulmonados). Ancestros de los peces pulmonados dieron origen a los tetrápodos.

Osteoblastos: Células activas formadoras de hueso. Sintetizan grandes cantidades de colágeno tipo I que forman la matriz ósea. Promueven la mineralización durante la formación y la remodelación de hueso.

Osteocitos: Células maduras del tejido óseo (las más abundantes) que derivan de osteoblastos. Se comunican entre sí y con osteoblastos por uniones comunicantes y tienen importantes funciones en la homeostasis del hueso.

Osteoclastos: Células multinucleadas responsables de la reabsorción ósea, derivan de las células del linaje monocitos-macrófagos.

Osteoide: Tejido orgánico de matriz no calcificada secretado por osteoblastos.

Parálogos: Son genes producidos a través de eventos de duplicación dentro de un genoma. Pueden evolucionar y mantener, adquirir o perder algunas o todas sus funciones (pseudogenización).

Pseudogen: Segmentos de ADN no funcionales que surgieron como copias de genes funcionales.

Pseudogenización: Proceso por el cual un gen puede adquirir una mutación espontánea que impida la transcripción o la traducción significativa, por ejemplo, debido a un cambio en el marco de lectura o en la introducción de un codón de stop prematuro.

Sintenia: Unión física de dos o más *loci* en la misma molécula de ADN. Sintenia conservada se refiere a los *loci* sinténicos que han mantenido su asociación física ancestral en dos o más linajes descendientes.

Teleosteos: Cualquier miembro de un grupo grande y extremadamente diverso de peces con aletas de raya.

Vertebrados: Subfilo de cordados que incluye los grupos de mamíferos, aves, reptiles, anfibios y peces. Suelen tener una columna vertebral ósea o cartilaginosa que reemplaza a la notocorda, una cabeza con un cerebro que surge como una parte ampliada de la médula nerviosa, y un esqueleto interno generalmente óseo (excepto en condriictos y lampreas).



RESUMEN

La familia del locus SCPP (Secretory Calcium-binding PhosphoProtein) codifica para componentes esenciales de matriz del tejido esquelético mineralizado en vertebrados. En el anfibio *Xenopus tropicalis*, el locus abarca 150 Kb y contiene los genes *Dmp1*, *Ibsp* y *Scppa2* que se expresan en hueso. Estos genes se originaron desde un ancestro común por duplicaciones en tándem. Durante el proceso evolutivo, cada uno de estos genes ha generado patrones de expresión divergentes. *Dmp1* se expresa en osteocitos mientras que *Ibsp* y *Scppa2* se expresan en osteoblastos y osteocitos.

El objetivo de este trabajo fue entender cómo los miembros del locus SCPP son regulados y por lo tanto pueden desplegar patrones de expresión divergentes. Se propone que dicha regulación involucra la acción de múltiples MCRs que contactan y coregulan promotores de miembros coexpresados en etapas específicas del proceso de diferenciación del linaje osteogénico.

El primer objetivo consistió en combinar hibridación *in situ* con tinción nuclear con DAPI usando pruebas exónicas (transcritos citoplasmáticos estables) e intrónicas (transcritos nucleares transcientes) y se demostró que *Ibsp* es un gen que se expresa activa y exclusivamente en osteoblastos y no en osteocitos. Los datos derivados de este trabajo revelaron un fuerte cambio regulatorio en la transición osteoblasto-osteocito. *Ibsp* es abruptamente apagado mientras *Dmp1* es rápidamente activado.

En todos los experimentos se incluyó también el gen *Sparc* debido a que comparte una historia evolutiva con el gen *Sparc1*, del cual derivaron los SCPP por duplicaciones en tándem y su expresión en hueso es similar a la del gen *Ibsp*.

Considerando que osteocitos directamente diferencian desde osteoblastos, la expresión osteocítica reportada previamente para *Ibsp* y *Sparc* se debe a la vida media larga de los RNA mensajeros y no a un proceso activo de transcripción.

Usando el enriquecimiento de marcas de cromatina asociadas a enhancers activos, se caracterizó la actividad transcripcional en osteoblastos de los MCRs putativos activos y ligados al locus SCPP y los MCRs ligados a *Sparc*. Se identificaron 2 enhancers que activan el promotor de *Ibsp*, uno proximal (fuerte) y uno distal (débil). Sorpresivamente, el promotor de *Scppa2* fue reprimido por dos MCRs, uno de ellos ligeramente activó el promotor de *Ibsp*. Ningún enhancer putativo activó el promotor de *Dmp1*, lo que era esperado dado que dicho promotor no es activo

en osteoblastos. Respecto a *Sparc*, al menos 3 enhancers fueron activos, dos proximales (fuertes) y un distal (débil).

Para evaluar si la arquitectura de la cromatina está involucrada en mantener unidos los miembros SCPP se utilizó el ensayo de Captura de Conformación de Cromosoma Circular (4C-seq) en osteoblastos primarios. Se identificaron las interacciones entre los promotores basales de los genes SCPP con diferentes regiones del genoma. También se demostraron las diferentes interacciones asociadas al promotor *Sparc*. Se demostró que los genes SCPP ácidos expresados en hueso se encuentran en un microdominio de interacciones frecuentes. El enhancer proximal que activó *Ibsp* tuvo una alta frecuencia de interacción con su promotor blanco. Se demostró también que los promotores *Ibsp* y *Dmp1* se contactan. Respecto a *Sparc*, los enhancers proximales de mayor actividad en estudios reporteros contactan a su promotor, el enhancer proximal además muestra cromatina abierta y es conservado en la evolución de vertebrados.

Los resultados de este trabajo permitieron entender cómo los miembros del locus SCPP se han mantenido en clúster, y cómo han adquirido expresión diferencial en el linaje osteogénico contribuyendo así al entendimiento de los mecanismos transcripcionales desplegados durante la transición osteoblasto/osteocito.



ABSTRACT

The vertebrate SCPP (secretory calcium-binding phosphoprotein) family locus encodes extracellular proteins involved in skeletal matrix mineralization. In the amphibian *Xenopus tropicalis* this locus spans 150kb and contains the neighboring genes *Dmp1*, *Ibsp* and *Scppa2*, known to have originated by tandem duplications from a common ancestor related to *Sparc1*. These genes have evolved divergent expression patterns and are expressed in osteocytes (*Dmp1*) or osteoblasts and osteocytes (*Ibsp* and *Scppa2*).

The aim of this work is to understand how SCPP member genes are regulated and therefore display divergent expression patterns. We examined multiple *cis* regulatory modules (MCRs) identified in the SCPP region that contact and coregulate promoters of coexpressed members at specific stages of the osteogenic lineage differentiation process.

To obtain a better understanding of the expression patterning of the SCPP genes we utilized DAPI staining along with *in situ* hybridization using exonic probes (stable cytoplasmic transcripts) and intronic probes (transient nuclear transcripts).

The *Sparc* gene is also examined in this study as a proxy for the ancestral condition of the SCPP. Indeed, *Sparc* is a *Sparc1* paralogue, and its bone expression is similar to the *Ibsp* gene.

We find that *Ibsp* and *Sparc* are only actively being transcribed in osteoblasts and not in osteocytes; however, stable transcripts are found in both osteoblasts and osteocytes. Considering that osteocytes directly differentiate from osteoblasts, the osteocytic expression is likely due to the long half-life of the mRNAs, and not to an active transcription process. Therefore, we identified a new regulatory switch at the osteoblast-to-osteocyte transition, in which the transcription of the *Ibsp* and *Sparc* is abruptly switched off while *Dmp1* is rapidly activated.

To better understand the processes involved in this transition, we looked for enrichment of chromatin labels associated with active enhancers. We identified putative *cis*-regulatory modules (CRMs) across commonly expressed genes. Using a novel reporter assay, we measured the transcriptional activity of the putative CRMs linked to the SCPP locus. The *Ibsp* promoter was strongly activated by one proximal CRM and weakly activated by one distal MCR. Surprisingly, the *Scppa2* promoter was repressed by two CRMs, one of which was weakly active on the *Ibsp* promoter.

No putative enhancer activated the *Dmp1* promoter, which was expected because this gene is not active in osteoblasts. The *Sparc* promoter was strongly activated by two proximal CRMs and weakly activated by one distal CRM.

To test if interactions with distal regulatory elements are involved in regulation of SCPP genes we looked at the 3D chromatin organization using a Circular Chromosome Conformation Capture assays (4C-seq) on primary osteoblasts. The interactions between the basal promoters of the SCPP genes with different regions of the genome were identified as well as the interactions associated with the *Sparc* promoter. Acidic SCPP genes expressed in bone were shown to be in a micro-domain of frequent interactions. The proximal enhancer that activated *Ibsp* had a high frequency of interaction with the *Ibsp* promoter. The *Ibsp* and *Dmp1* promoters were also shown to contact each other. Regarding *Sparc*, the proximal enhancers with the highest activity in reporter studies contact their promoter, the proximal enhancer also shows open chromatin and is conserved in the evolution of vertebrates.

The results of this work allowed us to understand how the members of the SCPP locus have remained in a cluster, and how they have acquired differential expression in the osteogenic lineage, thus contributing to the understanding of the transcriptional mechanisms displayed during the osteoblast / osteocyte transition.



1. INTRODUCCIÓN

1.1 Duplicaciones locales y familias de genes organizados en tándem

La expansión de familias génicas cumplió un rol crucial en la evolución de la complejidad anatómica, incluyendo el origen de nuevos tipos celulares y tejidos (Zhang & Cohn, 2008). El evento principal que originó esta expansión es la duplicación génica, que provee la materia prima que incrementa la posibilidad de adquirir nuevas funciones por selección natural. Copias de genes originados por duplicación que permanecen en el mismo genoma se denominan parálogos. Dentro de los mecanismos de duplicación génica se encuentran la poliploidización y el entrecruce desigual en la recombinación génica que resultan de errores en la meiosis (Magadum et al., 2013; Zhang, 2003).

Respecto al primer mecanismo, Susumo Ohno (Ohno, 1970) propuso un proceso de evolución por tetraploidía en la base de los vertebrados, una hipótesis que ha sido confirmada y que implicó dos rondas de duplicaciones del genoma completo conocidas como las “2R” (“two rounds” en inglés). Una ronda tuvo lugar en la raíz del linaje de los vertebrados, seguido por otra ronda que antecede a la separación entre agnatos (vertebrados sin mandíbula, e.g. lampreas) y gnatóstomos (vertebrados con mandíbula) (Figura 1). Las dos rondas de duplicaciones cuadruplicaron el contenido génico, generando copias de parálogos dispersos en el genoma (Dehal & Boore, 2005). Así fueron generando, por ejemplo, los cuatro loci de genes Hox que controlan la identidad de los segmentos corporales en el eje antero-posterior, así como el desarrollo de las extremidades (Duboule, 2007; Noordermeer et al., 2011). Sin embargo, frecuentemente los genes duplicados son silenciados por mutaciones degenerativas o no funcionales explicando por qué varias familias contienen menos de cuatro parálogos (Lynch & Conery, 2000). Por ejemplo, la familia SPARC perdió dos parálogos (Figura 1), dejando solo dos miembros: el gen SPARC (Secreted Protein Acidic and Rich in Cysteine) y el gen SPARC-L1 (SPARC-Like1) que están ubicados en cromosomas diferentes (Bertrand et al., 2013; Kawasaki et al., 2004).

Un segundo mecanismo de duplicaciones implica errores de recombinaciones que llevan a entrecruces desiguales los cuales producen secuencias repetidas en tándem. Dependiendo de la posición del entrecruce, la región duplicada puede contener la parte de un gen, el gen completo o varios genes (Zhang, 2003). Este tipo de duplicaciones locales representan un mecanismo clave para generar familias de genes ubicados en tándem en un mismo locus, como ha sido el

caso para la familia de genes SCPP (Secretary Calcium-binding Phosphoproteins) que provino de duplicaciones de SPARC-L1 (Figura 1); la familia de genes Hox, donde cada uno de los cuatro clusters, generados por las 2R, contienen aproximadamente 10 genes organizados en tándem (D. Duboule, 2007; Kuraku & Meyer, 2009); también los *loci* de los genes *Iroquois* (Tena et al., 2011) y los genes *Six* (Gomez-Marin et al., 2015) por mencionar algunos ejemplos.

Considerando que rearrreglos genómicos, inversiones, y translocaciones cromosómicas son muy frecuentes a escalas evolutivas, y que incluso estos eventos mutacionales han reordenado de manera sustancial el genoma de especies de mamíferos (Bourque et al., 2004; Murphy et al., 2005), las familias génicas en tándem invitan a plantearse las siguientes preguntas: **¿Cómo explicar que estos parálogos de vertebrados no han sido separados desde que fueron generados por duplicaciones locales casi 500 millones de años atrás?, en otras palabras ¿Qué mecanismo mantendrá unidos en un mismo locus a los miembros de familias génicas organizadas en tándem?** Posibles respuestas han emergido de estudios recientes de regulación transcripcional.

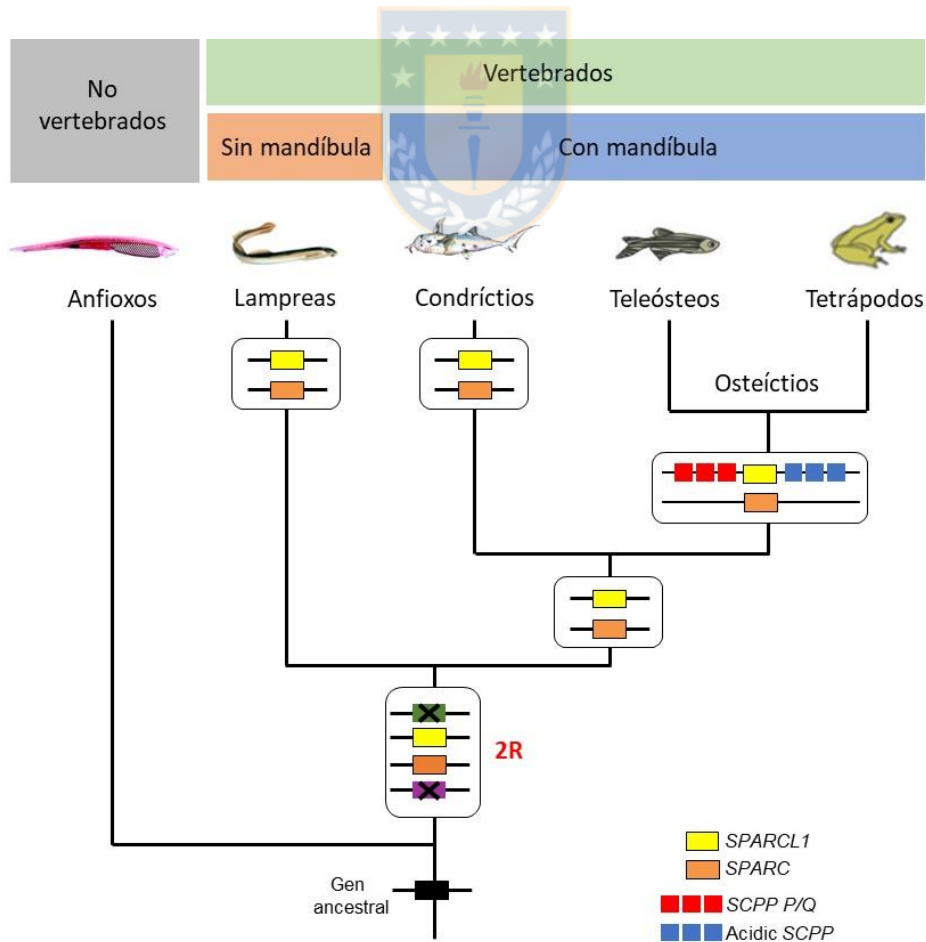


Figura 1: Origen del locus SCPP por eventos de duplicación génica

La duplicación de un gen ancestral por las dos rondas de duplicación del genoma completo (2R) originaron los genes SPARC y SPARCL1. Subsecuentes duplicaciones en tándem del gen SPARCL1 generaron la familia SCPP cuyos miembros sólo se han encontrado en osteóctios. Los genes SCPP se clasifican en dos categorías: un grupo codifica para proteínas ácidas (rectángulos azules) y el otro para las proteínas ricas en prolina y glutamina (P/Q) (rectángulos rojos). El locus SCPP está presente en todos los osteóctios y se asocia a grandes innovaciones evolutivas en vertebrados, siendo parte fundamental en procesos de mineralización en hueso, dentina y esmalte así como en la secreción de leche, lágrimas y saliva en mamíferos (Kawasaki & Amemiya, 2014). En el condrictio *C. Milii* se encontraron *Sparc* y *Sparc-like 1* (*Sparcl1*) en diferentes fragmentos cromosomales (*scaffolds*) que muestran sintenia conservada con *loci* ortólogos en vertebrados pero sin la presencia del clúster SCPP, lo mismo se ha observado en otros peces cartilagosos (condrictios) así como en lampreas (ciclóstomos). Esto sugiere que las duplicaciones en tándem de *Sparcl1* que dieron origen a los genes SCPP ocurrieron en el ancestro común de los osteóctios después de su ruptura con el linaje de los condrictios. La eliminación de un gen del clúster SCPP (*Spp1*) en *Zebrafish* por *CRISPR-Cas9* reduce la formación de hueso dando fuerte soporte a la hipótesis de que la ausencia de los genes SCPP en peces cartilagosos están relacionados con la naturaleza no osificada de su endoesqueleto (Venkatesh et al., 2014).

1.2 Regulación transcripcional

Todas las células de un organismo contienen prácticamente la misma secuencia de ADN, por lo que, para que las células se diferencien y realicen funciones específicas, los patrones espacio-temporales de expresión génica deben estar finamente regulados. Esto se consigue principalmente mediante la regulación de la producción de transcritos. La regulación transcripcional se consigue a través de redes reguladoras complejas que ejercen el control a través de varios mecanismos, como la modulación de la estructura de la cromatina, la disponibilidad de proteínas, el inicio de la transcripción, la elongación y el *splicing* del ARNm (Shlyueva et al., 2014).

En la regulación transcripcional un paso clave es la modulación de la fase de iniciación, en la que intervienen diferentes elementos del ADN, modificaciones en las marcas de cromatina y el reclutamiento de factores de transcripción generales y específicos de secuencia en sus sitios blanco en el ADN. Las secuencias de ADN implicadas son el promotor, adyacente al sitio de inicio de la transcripción (TSS), y elementos proximales y/o distales como enhancers, silenciadores e insuladores (Phillips & Corces, 2009).

El promotor basal es el lugar donde se ensambla el complejo de preiniciación antes del inicio de la transcripción. Dicho complejo está formado por la ARN polimerasa II (RNAPII), factores de transcripción generales (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF y TFIIH) y el Mediador. El Mediador es un gran complejo proteico que interactúa con factores de transcripción de secuencia

específica, transmitiendo sus señales a la polimerasa II, modulando su función, por ejemplo, fosforilándola, lo que conduce a la liberación del complejo de preiniciación. El Mediador participa en la regulación transcripcional alterando la organización de la cromatina y modulando el bucle enhancer-promotor y el inicio y la elongación de la transcripción (Kagey et al., 2010).

Inicialmente, los promotores fueron caracterizados por la presencia de la caja TATA (localizada 25-35 pb antes del sitio de inicio de la transcripción del gen), la cual esta presente en una minoría de casos y asociada con genes tejido-específicos y de alta conservación entre especies. Otra característica es la presencia de islas CpG, presente en la mayoría de promotores y asociada a procesos de evolución rápida, debido a su plasticidad y capacidad de regulación transcripcional ligada a procesos de control epigenético (Carninci et al., 2006). El promotor basal por sí sólo no es suficiente para una eficiente transcripción por lo que se requiere de los **módulos cis-reguladores** (MCRs), que modulan la actividad del promotor de sus genes blanco (Zabidi & Stark, 2016), y por lo tanto, los niveles de transcripción.

La regulación transcripcional requiere la precisa orquestación de un complejo grupo de interacciones entre proteínas (los factores *trans*) y segmentos no codificantes de ADN (de tamaño ~0.1-1.5 kb) que actúan modulando la transcripción, los (MCRs). Las secuencias de los MCRs contienen cortos dominios de ADN que actúan como sitios de unión para factores de transcripción (FT) y junto con proteínas coactivadoras o corepresoras, forman un complejo que promueve o impide el reclutamiento de la ARN polimerasa a promotores de genes específicos (Shlyueva et al., 2014). Las secuencias de unión a factores de transcripción son típicamente de un tamaño menor a 30 pares de bases (pb). Los patrones de secuencia conservados de una colección de sitios de unión similares se conocen como **motivos** y son elementos esenciales que contribuyen al entendimiento de los mecanismos de regulación de la expresión génica (Inukai et al., 2017). Los MCRs pueden encontrarse en regiones intergénicas e intrónicas en el genoma; pueden estar inmediatamente corriente arriba del promotor basal (proximales: <10 kb) o a grandes distancias (distales: usualmente >10 kb) y son llevados a la proximidad de sus promotores blanco mediante bucles de cromatina, modificando así la expresión génica. Un ejemplo clásico es un MCR distal intrónico (~1 Mb) en el gen *Lmbr1* que regula la expresión del gen *Shh* en el contexto del desarrollo de extremidades y cuya desregulación se asocia a polidactilia preaxial (Lettice et al., 2003). La unión de los FT a los MCRs puede iniciar, aumentar o suprimir la transcripción. De acuerdo a su función reguladora, los MCRs pueden ser categorizados como promotores, enhancers, silenciadores e insuladores (Shlyueva et al., 2014).

El rol de los **enhancers** es aumentar la actividad transcripcional de un promotor. La localización de estos elementos respecto a sus genes blanco es variable, pueden ubicarse río arriba, río abajo, en intrones de los mismos genes regulados o en diferentes locus. Los mecanismos moleculares que subyacen a los enhancers son ampliamente estudiados y tienen una gran relevancia en contextos de desarrollo, evolución y en patologías (Franke et al., 2016; Lupiáñez et al., 2015). La vasta mayoría de enhancers estudiados hasta el momento tienen una relevancia funcional debido a su actividad espacio-temporal por lo que patrones de expresión génica han sido ligados a enhancers específicos, destacando así su naturaleza modular. Sus propiedades son independientes del contexto de secuencias en el que se encuentran, por lo que ensayos de expresión génica utilizando genes reporteros ha sido fundamental en el descubrimiento de su actividad debido a la facilidad de su recapitulación funcional en un contexto de cromatina.

Por otra parte, los **silenciadores** inhiben la actividad transcripcional y los **insuladores** bloquean las interacciones promotor-enhancer, sirviendo como barreras contra la propagación de los efectos silenciadores de la heterocromatina. Un ejemplo bien conocido de insulador es el sitio de unión al factor CCCTC (CTCF) que previene el entrecruce entre regiones genómicas activas e inactivas y está implicado en diversos roles de regulación génica que van desde la activación/represión de promotores dependientes de contexto, el bloqueo de enhancers o barrera de insulación, el silenciamiento de respuesta hormonal, el *imprinting* genómico y finalmente, su relevancia en las interacciones de cromatina a grandes distancias (Phillips & Corces, 2009). Los mecanismos moleculares por los cuales MCRs y promotores interactúan a distancia y su relevancia en la expresión génica son actualmente un campo de investigación activo.

1.3 Regulación epigenética de la expresión génica

La actividad de los MCRs está correlacionada con propiedades de la cromatina. Por ejemplo, los enhancers activos son típicamente libres de nucleosomas (unidades estructurales de la cromatina), lo que permite accesibilidad para los FT. Además, los nucleosomas en la vecindad de enhancers activos típicamente contienen histonas con determinadas modificaciones postraduccionales. Estas “marcas epigenéticas” modifican la estructura de la cromatina y por lo tanto la accesibilidad al ADN, lo que posibilita otro aspecto de regulación génica. Las histonas forman octámeros que contienen dos copias de cada una de las cuatro proteínas (H2A, H2B, H3 y H4). Las modificaciones se llevan a cabo mayoritariamente en el N-terminal de las colas de histonas y pueden involucrar metilación, acetilación, ubiquitinación y sumoilación, entre otras.

Los patrones de modificación sugieren que existe un “código de histonas” que regulan la organización y función de la cromatina. De hecho, marcas individuales de histonas han hecho distinguible la eucromatina (transcripción activa) de la heterocromatina (transcripción silenciada). La primera es caracterizada por la metilación de la lisina 4 de la histona 3 (H3K4me) mientras que la segunda por la metilación de la lisina 9 de la histona 3 (H3K9me) (Noma et al., 2001). Además, modificaciones particulares de histonas son asociadas con diferentes estados transcripcionales. Por ejemplo, la mono-metilación de la lisina 4 de la histona 3 (H3K4me1) y la acetilación de la lisina 27 de la histona 3 (H3K27ac) son modificaciones específicas que permiten detectar enhancers y promotores activos. Los enhancers inactivos pueden ser silenciados por diferentes mecanismos, dos clásicos son la triple metilación de la lisina 27 de la histona 3 (H3K27me3), que es reconocida por el complejo Polycomb, y la actividad de FT represivos (Hoffman et al., 2013).

Estas marcas epigenéticas han permitido entender los complejos patrones de expresión génica que resultan de la acción sumada de diferentes enhancers con actividades celulares o tejido tipo-específicas. Dichas marcas han sido bien categorizadas en levadura, mamífero y *Drosophila* principalmente mediante estudios de secuenciación genómica (ChIP-seq), sin embargo, aún hay mucho por descubrir en otros modelos. Por ejemplo en el anfibio *Xenopus* se han hecho importantes avances para entender la funcionalidad de las marcas de histonas a nivel del desarrollo temprano o en tejidos específicos (Hontelez et al., 2015), sin embargo, hay falta de información en estadios post-larvarios donde órganos complejos y tejidos mineralizados emergen.

1.4 Dominios de Asociación Topológica (TADs)

El genoma de las células eucariotas debe sufrir un fino empaquetamiento a nivel nuclear que es logrado por un conjunto de complejos moleculares hechos de proteínas y ácidos nucleicos dispuestos en una forma extraordinariamente ordenada que permiten de manera dinámica que miles de regiones del genoma se vuelvan accesibles a la maquinaria de transcripción de genes de forma orquestada en tiempo y espacio. Esta característica ha llevado a la investigación de las propiedades funcionales de la topología de los genomas. Debido a que se ha corroborado la existencia de MCRs funcionales en posiciones genómicas alejadas linealmente de sus genes blanco, se ha investigado si estos elementos actúan a distancia mediante contactos con los

promotores de sus genes blanco a través de estructuras en bucle o dominios de cromatina permitiendo una proximidad en el espacio nuclear que se acopla con la maquinaria de transcripción para generar la especificidad requerida de la expresión génica. Observaciones microscópicas del núcleo celular utilizando sondas de ADN fluorescentes han hecho posible determinar regiones en las que segmentos cromosomales alejadas linealmente pueden aproximarse e interactuar en regiones nucleares específicas (Giorgetti & Heard, 2016).

Interacciones a largas distancias entre *loci* específicos han sido comprobadas principalmente mediante la técnica de captura de conformación de cromosoma (3C) y sus variantes (4C, 5C y HiC), que consisten en la caracterización de frecuencias de interacción entre regiones del genoma. La diferencia entre estas metodologías radica, principalmente, en la cantidad de contactos entre *loci* que pueden evaluarse en un solo experimento (ver sección 6 para fundamento teórico de 4C). En 2009 surgió una variante de estas técnicas denominada **HiC** que corroboró un nivel de organización adicional de los genomas que se caracteriza por la segregación espacial de cromatina abierta y cerrada (compartimentos A y B) y sugirió que la conformación tridimensional de la cromatina permite que regiones alejadas en el genoma pueden tener una proximidad de elementos funcionales (Lieberman-Aiden et al., 2009). En 2012 se propuso que la cromatina de mamíferos está organizada en Dominios de Asociación Topológica (TADs por sus siglas en inglés), definidos como bloques genómicos a escalas de megabases dentro de los cuales secuencias de ADN exhiben significativamente altas frecuencias de interacción con otras secuencias dentro del dominio pero no con regiones externas (Dixon et al., 2012; Nora et al., 2012), estas observaciones se extendieron rápidamente a otros organismos y tipos celulares (Hou et al., 2012; Phillips-Cremins et al., 2013; Sexton et al., 2012) y recientemente la técnica de HiC se aplicó por primera vez en *Xt*, donde se corroboró la naturaleza de los TADs y se observó, entre otras cosas, su dinámica en la embriogénesis (Niu et al., 2020). Los TADs se descubrieron con bajo nivel de resolución (1 Mb) y en un principio se propusieron como elementos estables a lo largo de distintos tipos celulares y altamente conservados entre ratón y humano (Dixon et al., 2012). Sorprendentemente, con la ultra-resolución actual (1-4 kb), se ha demostrado que los TADs tienen dinámicas específicas de tipo celular, es decir, los contactos pueden variar en el proceso de diferenciación (Beagan & Phillips-Cremins, 2020).

Los TADs se encuentran demarcados por fronteras o límites proteicos en los que convergen principalmente las proteínas CTCF y cohesinas (proteína de la familia de complejos proteicos que mantienen la estructura de los cromosomas). Entre los límites existen dominios de cromatina en escalas sub-megabase denominados subTADs, sólo una porción de ellos fue observada en los primeros estudios de HiC, pero actualmente se ha demostrado que los

subTADs “anidados” se asemejan a TADs demarcados por claros límites pero con débiles marcas de insulación (Beagan & Phillips-Cremins, 2020; Hansen, 2020). Los TADs se determinan por la significativa frecuencia de interacción. Se les ha denominado también “dominios de bucles” y su tamaño puede variar desde algunos kilobases hasta megabase. Más del 60% de TADs en mamíferos contienen dominios de unión a CTCF con una orientación de motivo convergente que lleva a la formación de bucles (Beagan & Phillips-Cremins, 2020). De acuerdo al modelo de *extrusión de bucle*, la cohesina, después de ser cargada sobre la cromatina por el factor de carga NIPBL, extrae la cromatina a través de una estructura con forma de anillo, lo que resulta en una forma de bucle en crecimiento que se detiene cuando se encuentra con los sitios ocupados por el factor CTCF (Figura 2A). Células en ausencia de cohesina muestran una severa afectación de generación de bucles y células deficientes del factor de descarga de la cohesina (WAPL) incrementan el tiempo de residencia de la proteína sobre la cromatina coincidiendo con el aumento del tamaño de los bucles y por lo tanto un mayor número de TADs o subTADs mientras que la inactivación del complejo de carga de la cohesina (NIPBL) lleva al acortamiento de los bucles (Schoenfelder & Fraser, 2019).

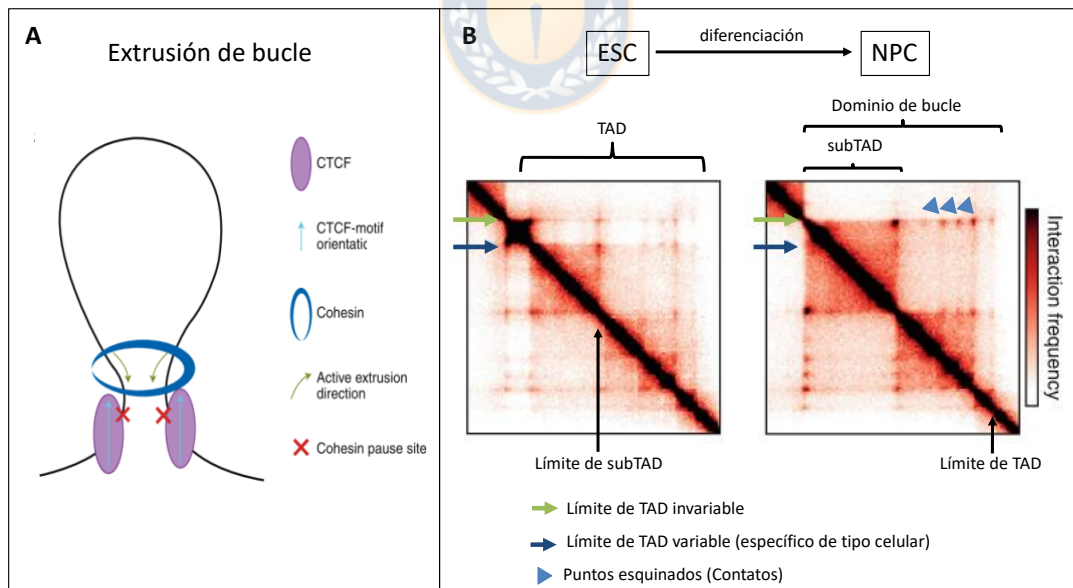


Figura 2: **Mecanismo de formación de TADs**

En A el modelo de extrusión de bucle muestra que la translocación de la cohesina extruye el ADN en una forma dependiente de ATP generando interacciones de bucle distales en una zona que contiene CTCF en dirección convergente. Estas interacciones forman la base topológica de los dominios de bucle TAD y subTADs. En B se muestran los mapas de frecuencia de contactos de datos HiC de alta resolución

derivados de células troncales embrionarias (ESC) y células neuroprogenitoras (NPC). Un TAD contiene dominios de bucle evidentes y subTADs anidados. Los triángulos azules representan las esquinas de un dominio de cromatina que forma bucles evidentes en una determinada región genómica. Las flechas verdes señalan un límite de TAD “fuerte” (invariable de tipo celular) mientras que las flechas azules señalan un límite “dinámico” de tipo celular.

La funcionalidad de los TADs es un tema vigente que ha generado muchas preguntas de investigación que van desde sus propiedades mecánicas hasta su relevancia en determinar y dirigir la transcripción a través de los intrínsecos contactos entre MCRs. Actualmente se ha desarrollado una variante de HiC denominada Micro-C, la cuál permite capturar relevantes interacciones enhancer-promotor y “micro-dominios” (escala de Kbs) que eran inaccesibles por HiC (Hsieh et al., 2020; Krietenstein et al., 2020) y se ha logrado confirmar, entre otras cosas, que la diferenciación celular o el estado regulatorio está asociado a cambios generalizados en la organización de bucles de los TADs (Figura 2B).

Una primera aproximación para averiguar si la estructura de la cromatina 3D tiene un papel preponderante en la regulación génica fue eliminar o modificar los bordes de los TADs (CTCF y cohesinas). Varios estudios, en loci específicos, mostraron que la disrupción de los límites causa contactos ectópicos inter-dominio entre enhancers y promotores erróneos, afectando la expresión génica y llevando a irregularidades en el desarrollo, cáncer, desórdenes neurológicos y malformaciones (Sivakumar et al., 2019). No obstante, cuando se afectaron de manera global los límites de TADs mediante la pérdida de la cohesina por un sistema inducible por auxina que permite la eliminación específica de proteínas por ubiquitinación y subsecuente degradación proteosomal, se observaron sólo modestos efectos de la expresión génica, aún cuando los dominios de bucle fueron eliminados (Rao et al., 2017). Además, de manera específica, la eliminación del TAD en el que se encuentran los contactos enhancer-promotor que regulan la expresión del gen *Shh* (expuesto anteriormente), no llevó a efectos detectables de expresión durante el desarrollo, excepto cuando los enhancers fueron eliminados (Williamson et al., 2019). Un efecto similar fue observado cuando se eliminaron los bordes del TAD que contiene el locus de *Sox9-Kcnj2 in vivo* en ratón en donde incluso la fusión de dos TADs tuvo apenas efectos moderados en la expresión génica. Sin embargo, desregulación y cambios fenotípicos fueron observados cuando se invirtieron los límites de TADs ya que se perdieron contactos entre los promotores de *Sox9* y *Kcnj2* y sus enhancers (Despang et al., 2019). Dado que las inversiones de los límites entre TADs tienen efectos pronunciados en la expresión génica, los rearrreglos cromosomales que afectan estos límites en vertebrados son hipotetizados por tener efectos severos en la transcripción (Ibrahim & Mundlos, 2020). Cada vez es más evidente que los dominios de asociación topológica se asemejan a un andamio flexible que optimiza las interacciones

enhancer-promotor. Hay un campo de investigación activo sobre la naturaleza dinámica de la conformación de la cromatina y su función en la regulación génica.

1.5 Arquitectura regulatoria en vertebrados

Enhancers distales son los más abundantes en los vertebrados y son críticos para la precisa expresión espacio-temporal de genes que regulan el desarrollo animal (Schoenfelder & Fraser, 2019). A escala evolutiva, traslocaciones cromosómicas e inversiones tienden a modificar la sintenia (orden de los genes en un cromosoma) entre especies. Sin embargo, existen excepciones a este proceso mutacional, y algunos genes permanecen adyacentes unos de otros a pesar de millones de años de divergencia (Simakov, 2020). Se ha demostrado que dichos genes comparten regulación de la expresión génica, así y de forma coordinada, su expresión depende de uno o más enhancers compartidos, explicando porque permanecen en un mismo locus (Tena et al., 2011). Efectivamente, los genes que comparten elementos reguladores conservados tienen una fuerte presión selectiva a favor de la mantención de la organización en dominios genómicos asociados al desarrollo en vertebrados (Woolfe et al., 2004). Basado en estas observaciones, se generó la hipótesis de que elementos reguladores alrededor de estos *loci* inducen la restricción evolutiva que mantiene la estabilidad de genes sinténicos, lo que explicaría la ausencia de “puntos de ruptura” en grandes regiones cromosomales (Kikuta et al., 2007; Vavouri et al., 2006). Mediante análisis de conservación de MCRs y clonamiento de estas regiones en vectores reporteros se demostró que los genomas de vertebrados contienen MCRs distales distribuidos en grandes áreas dentro y alrededor de sus genes blanco y rodeados filogenética y funcionalmente de genes vecinos (“bystander”), formando regiones de sintenia conservada. Estos segmentos fueron denominados bloques de regulación genómica (GRBs) y se propusieron como un mecanismo que explicaría la conservación de sintenia en el genoma de vertebrados. El *bystander* se define como un gen que no está específicamente bajo el control de los enhancers que definen el GRB en el cual está localizado, es decir, son físicamente presentes en la región genómica donde los eventos regulatorios toman lugar pero no participan en su regulación (Kikuta et al., 2007). Los GRBs pueden encontrarse incluso en grandes regiones del genoma (hasta 250 kb en promedio) generando así una complejidad regulatoria en dichos loci debido a que un gen puede ser contactado por varios MCRs, o viceversa un MCR puede contactar a diferentes genes estableciendo una importante presión para mantener los reguladores y sus genes blanco juntos, por lo tanto, el costo de romper las relaciones físicas entre regiones reguladoras distales de sus genes blanco puede ser tan alto que los re-arreglos

son raramente fijados en el proceso evolutivo (Irimia et al., 2013; Vavouri & Lehner, 2009). Copias únicas de GRBs suelen ser protegidas de la ruptura cromosómica, sin embargo, en casos de duplicación de GRBs (por ejemplo en teleósteos) los genes bystander funcionalmente no relacionados al dominio de regulación genómica han sido borrados por evolución neutral, un fenómeno predicho por el modelo de duplicación-degeneración-complementación (Kikuta et al., 2007).

Los GRBs coinciden con los TADs (Harmston et al., 2017), lo que era esperado si se considera que la definición de TADs implica regiones genómicas donde los contactos entre promotores y enhancers son frecuentes. La correspondencia entre los elementos no codificantes reguladores que se distribuyen a lo largo de los GRBs y los TADs sugiere que la arquitectura de la cromatina es conservada. Existen ejemplos claros de unión física de dos o más *loci* en un mismo fragmento de ADN donde elementos no codificantes conservados y sus genes blanco mantienen la sintenia en grandes periodos evolutivos como resultado de la mantención de los elementos reguladores en *cis* con el gen o genes blanco (Wong et al., 2020). Debido a que las regiones reguladoras podrían encontrarse a grandes distancias de sus genes blanco se propuso que debían existir interacciones entre los promotores basales con sus enhancers, lo cual necesariamente implica una proximidad física en el núcleo. Así, interacciones entre elementos reguladores localizados en regiones intergénicas o en intrones distales con sus promotores han llevado a preguntarse los posibles mecanismos por los cuales se lleva a cabo esta compleja regulación génica. Sólo algunos clústers de genes corregulados por varios enhancers en grandes regiones genómicas de macrosintenia conservada y que codifican para factores de transcripción han sido estudiados desde el punto de vista de la arquitectura regulatoria.

Un ejemplo paradigmático es el caso de los genes Hox que han permanecido asociados en clúster desde que se originaron e incluyen “paisajes regulatorios” en donde las interacciones entre promotores y enhancers son fundamentales para su control espacio-temporal en el desarrollo. Análisis puntuales de arquitectura de la cromatina de los loci *HoxD* y *HoxA* han revelado la existencia de dos TADs de tamaño aproximado a 1Mb que abarcan precisamente los paisajes regulatorios adyacentes a los clúster, estos últimos flanqueados en el extremo 5' y 3' respectivamente por múltiples enhancers distales que regulan su expresión durante el desarrollo de extremidades en vertebrados (Lonfat & Duboule, 2015). La modificación de este “archipiélago” regulatorio puede directamente impactar la morfología de los dígitos, sugiriendo que cambios en elementos reguladores proveen una base para la diversidad en las formas y números de dígitos en la evolución de los tetrápodos así como en diferentes síndromes en humanos (Montavon et al., 2011; Noordermeer et al., 2011) (Figura 3A). Un elegante mecanismo para explicar la

dinámica del archipiélago regulatorio de los genes *HoxD* durante el desarrollo de extremidades fue propuesto por el grupo de Denis Duboule en 2013: en una fase temprana, los enhancers dentro del TAD 3' (T-Dom) del clúster *HoxD* activan y controlan la expresión de los genes *Hox* del 3', los cuales están involucrados en la formación de estructuras proximales (antebrazo), mientras que los enhancers del TAD 5' (C-Dom), en una fase tardía, promueven la expresión de los genes *Hox* 5', en regiones distales de las extremidades (manos y dígitos). Mientras progresa el desarrollo, los genes de la parte central del clúster transitan del 3' al 5' TAD cambiando su paisaje regulatorio y por lo tanto su expresión del territorio proximal al distal de las extremidades generando una importante transición regulatoria (Andrey et al., 2013) (Figura 3B). ¿Cómo se controla la regulación temporal de esta transición? Un estudio relativamente reciente mostró que las proteínas HOX13 son críticas ya que activan un mecanismo represivo global en el 3' TAD mientras que al mismo tiempo, los contactos en el 5' promueven su actividad generando un bucle de retroalimentación positivo (Maeso et al., 2017). Interesantemente, parte del 5' TAD es también usado en el desarrollo de genitales externos. Ambas estructuras hacen uso de algunos enhancers específicos localizados dentro de la misma arquitectura regulatoria sugiriendo un ambiente de diversidad funcional de *loci* del desarrollo por secuestrar la maquinaria reguladora de ciertos enhancers de un tipo celular a otro (Amândio et al., 2020). De forma relevante, en especies donde el tamaño de genoma se ha reducido drásticamente y donde las interacciones regulatorias a grandes distancias no son preponderantes, como es el caso en los tunicados (un grupo de invertebrados del phylum de los cordados), el locus *Hox* ha sido totalmente "atomizado", es decir, los genes *Hox* se encuentran esparcidos en el genoma y separados unos de otros (Denis Duboule, 2007).

Otro ejemplo paradigmático de GRBs es la familia de genes *Iroquois*, que están organizados en clústeres dentro de los cuales existen enhancers distales compartidos (interactúan con más de un promotor *Irx*) y son altamente conservados entre vertebrados. Estos clústeres están formados por 3 genes que presentan patrones de expresión diferencial. Los dos primeros genes de un clúster se expresan en el tejido neural, riñón y vesícula óptica, mientras que la expresión del tercer gen se restringe al tejido neural. ¿Cómo se lleva a cabo este patrón de expresión divergente? Gómez-Skarmeta y colaboradores en 2011 encontraron que existen MCRs que preferentemente contactan a genes *Irx* co-expresados (Figura 3C). Los autores demostraron la existencia de una arquitectura cromatínica o estructura 3D donde los promotores de los primeros dos genes de cada clúster forman un bucle que los acerca y los lleva a una misma zona regulatoria que probablemente facilita su corregulación, reduciendo al mismo tiempo, el acceso al promotor de un tercer gen que presenta una regulación y expresión

diferencial (Figura 3D). Esta disposición coloca a los dos primeros genes en un mismo paisaje regulatorio, facilitando así su coregulación, lo que significa que dos promotores de genes que presentan un mismo patrón de expresión pueden ser contactados por enhancers compartidos. Sin embargo, la misma arquitectura regulatoria 3D reduce el acceso al promotor del tercer gen del complejo generando un contexto regulatorio diferencial para este último. Si este modelo es correcto, la disrupción del clúster por inversiones o translocaciones cromosómicas podría fuertemente afectar la expresión de los genes que se encuentran separados de sus enhancers, con consecuencias nefastas sobre la función celular y el individuo mutante. Así, la presencia de enhancers compartidos actuaría como un “candado molecular” que impide la separación de los genes en tándem, una restricción evolutiva que mantendría la asociación de los genes *Irx* en clúster (Tena et al., 2011). Otros GRBs de microsintenia conservada a través de los metazoos y que han sido estudiados desde una perspectiva de la arquitectura regulatoria son los genes *Pax1/9* y *FoxA* (Wang et al., 2007), *Hedgehog* (Irimia et al., 2012), *Six* (Gomez-Marin et al., 2015) y *Dlx* (Sumiyama & Tanave, 2020).

Un estudio reciente de evolución de GRBs en el genoma humano ha generado una lista de 745 genes con funciones de factores de transcripción importantes en el desarrollo pertenecientes a 363 familias que ya estaban presentes en el último ancestro común de los cordados. Al menos 156 GRBs putativos son conservados después de las dos rondas de duplicación del genoma completo en vertebrados y mantenidos en humanos con al menos una asociación de microsintenia conservada entre genes del desarrollo y sus byestander (31 de estos genes están unidos a dos o más genes bystander) (Touceda-Suárez et al., 2020). En esta lista se encuentra la relación sinténica clásica del gen *Runx2* con *Supt3h* que será estudiada en la siguiente sección. En dicho estudio se contemplaron únicamente genes que codifican para FT, sin embargo, no aparecen aquellos que duplicaron posteriormente en tándem como los SCPP, los cuáles sólo se originaron en la base de los osteíctios. Otros ejemplos relevantes de microsintenia conservada en tejidos mineralizados son la familia de genes *SPARC-L1/SPARC-L2* entre condriictios (clase de peces de esqueleto cartilaginoso) y osteíctios (comúnmente llamados organismos de esqueleto óseo) (Enault et al., 2018) y los genes del colágeno *10a1* que duplicaron localmente en tándem en el clado de los condriictios y que mantienen sintenia conservada con osteíctios (Debiais-Thibaud et al., 2019).

En conclusión, los trabajos desarrollados hasta el momento en arquitectura regulatoria subrayan la importancia de los MCRs en imponer la restricción de la arquitectura genómica, incluso a través de grandes distancias filogenéticas, lo que impacta funcionalmente en la diferenciación celular y evolución de la novedad en tetrápodos.

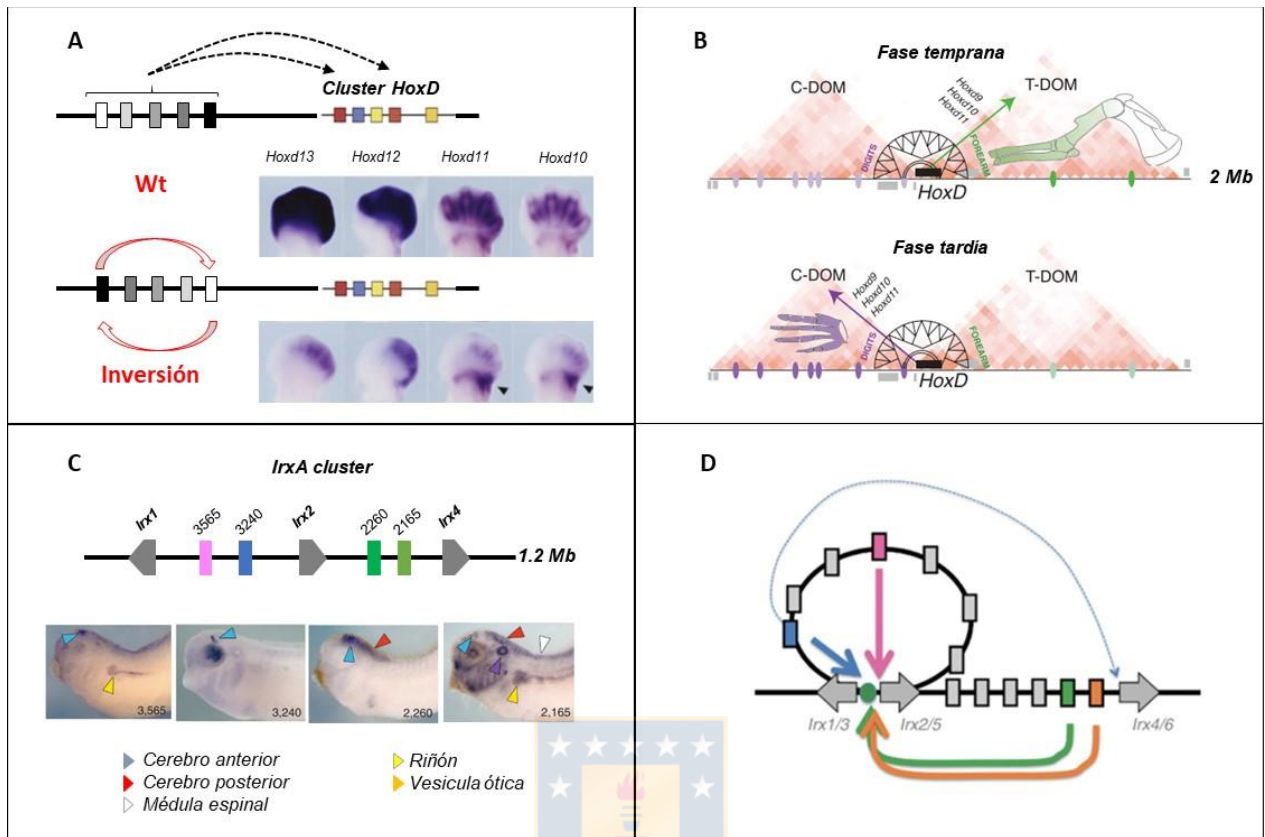


Figura 3: Ejemplos de arquitectura regulatoria

En A se esquematiza un archipiélago regulatorio que dirige la transcripción de los genes *HoxD* en el desarrollo de los dígitos en ratón. La inversión de estos elementos reguladores utilizando el sistema de recombinación *Cre-Lox* reubica la posición de los enhancers. Los animales homocigotos con la inversión tienen los dígitos más cortos y las falanges perdidas o fundidas, un fenotipo muy similar a la eliminación de los genes *Hoxd8* y *Hoxd13* (no mostrado). La expresión distal de los genes *Hoxd13-Hoxd10* fue notablemente reducida mientras *Hoxd11* y *Hoxd10* fueron todavía expresados en el dominio de la extremidad proximal (puntas de flecha negras) indicando que su efecto regulatorio no afectó directamente la fase de expresión temprana (Montavon et al., 2011). En B se muestra la regulación de la expresión colineal de los genes *Hoxd* en el desarrollo de extremidades. El clúster de genes está flanqueado por dos dominios de asociación topológica que contienen capacidad de regulación diferencial. El dominio telomérico (T-DOM) regula la transcripción temprana en el brazo y antebrazo. El dominio centromérico (C-DOM) subsecuentemente dirige la expresión de patrones de las manos. Durante el desarrollo de extremidades un *switch* ocurre entre los archipiélagos regulatorios llevando a que genes específicos del cluster *Hoxd* oscilen de un dominio regulatorio al otro a través de cambios conformacionales de la cromatina. Esta organización modular crea un dominio intermedio que forma la muñeca (Andrey et al., 2013). En C se muestra el GRB de los genes *Iroquois A* que incluye los genes *Irx1-Irx3* así como 4 enhancers conservados en vertebrados representados con diferentes números y cajas de colores. Estas regiones fueron amplificadas de ADN de *Xt* y clonadas para ensayos de transgénesis *in vivo*. Los enhancers condujeron la expresión de reporteros GFP en subdominios de patrones endógenos de los genes *IrxA* en cerebro medio y posterior, médula espinal, riñón y vesícula ótica (detección por hibridación *in situ* del ARN de GFP) (Tena et al., 2011). Estas regiones condujeron la expresión en tejidos similares

en ratón y pez cebra (no mostrado). En D el modelo de arquitectura 3D de los clústeres /rx. El esquema muestra un bucle de ADN con los miembros del locus (flechas grises). La proximidad de los promotores de los primeros dos genes facilitada por CTCF (círculo verde) lo que da acceso similar de los enhancers (cajas azul y rosa) a dichos promotores. Un enhancer (caja azul, línea punteada) es de restringido acceso a un tercer promotor y no puede actuar eficientemente (línea azul discontinua). Las cajas grises representan otros enhancers identificados en el locus. Todos los enhancers representan posiciones relativas analizadas por ensayos de conformación de la cromatina 3C (Tena et al., 2011).

1.6 Patrones de expresión y función de los miembros SCPP y SPARC

La familia de fosfoproteínas de unión a calcio secretado (SCPP por sus siglas en inglés) es encontrada en tejidos mineralizados de osteóctios y está involucrada en la unión de calcio y fosfato. Su función principal en tejidos mineralizados es nuclear los cristales de hidroxiapatita. Una alta tasa de ganancia, pérdida y divergencia de genes que producen esta familia ha dramáticamente modificado el repertorio de genes en distintos linajes de osteóctios (Kawasaki, 2009). Por ejemplo, en mamíferos, importantes innovaciones de la familia por duplicaciones en tándem generaron proteínas que se secretan también en fluidos como la leche, saliva, lágrimas y orina. En aves, miembros del clúster SCPP se secretan en la producción de la cáscara de huevo, cuya capa interna sirve como reservorio de calcio para el esqueleto del embrión en desarrollo. Interesantemente, la preservación de agua y calcio por la cáscara de huevo en las aves es funcionalmente paralela con el suplemento de estos materiales en la leche materna, dos importantes características asociadas a la innovación de los miembros del locus SCPP. Además, mientras las aves ancestrales tempranas tuvieron una dentadura simple, las aves modernas tienen completamente perdidos sus dientes, lo que está fuertemente asociado a la *pseudogenización* de miembros del locus involucrados en la mineralización del esmalte (Kawasaki & Weiss, 2006).

Los genes SCPP se expresan principalmente en dientes y hueso (Kawasaki et al., 2004). Los dientes están formados por esmalte y dentina. El esmalte es un tejido hipermineralizado que cubre la superficie del diente y es secretado por ameloblastos, células de origen ectodermal. El cuerpo interno del diente se compone de dentina, secretada por odontoblastos de origen mesenquimal (Chinsembu, 2012; Hall, 2005; Kawashima & Okiji, 2016). Por otra parte, el desarrollo del hueso, u osteogénesis, involucra al menos tres tipos celulares que pertenecen a un mismo linaje: células mesenquimáticas u osteocondroprogenitoras, osteoblastos y osteocitos (Hall, 2005). Una vez que los progenitores reciben señales osteogénicas, principalmente de factores secretados por condrocitos hipertróficos, se comprometen a osteoblastos precursores, los cuales maduran en osteoblastos que depositan la matriz ósea (Hojo & Ohba, 2020).

Finalmente, los osteoblastos quedan embebidos en su propia matriz y diferencian a osteocitos, células que tendrán un papel preponderante en la homeostasis del hueso (Robling & Bonewald, 2020).

Los genes SCPP se encuentran organizados en clúster en osteóctios (Figura 4) y se clasifican en dos subclases que codifican para dos subfamilias de proteínas. **Las ricas en prolina y glutamina (P/Q)**, como amelogenina (AMEL), ameloblastina (AMBN) y enamelinina (ENAM), son expresadas en ameloblastos y requeridas para la esmaltogénesis, mientras que las **proteínas ácidas**, denominadas también **SIBLING** (Small Integrin Binding Ligand N-Glycosylated), son necesarias para la correcta mineralización de hueso y dentina (Fisher et al., 2001; Kawasaki, 2009). Aunque estas proteínas son muy divergentes en su secuencia primaria, comparten características estructurales similares: i) la modificación post-traducciona por fosforilación es la más importante en sus funciones de matriz extracelular, ii) poseen una secuencia consenso de unión a integrina RGD (Arg-Gly-Asp) la cual es un sitio de anclaje celular que facilita la adhesión a la matriz y iii) contienen un péptido ASARM, el cual puede ser hidrolizado por una proteína PHEX, un inhibidor de la mineralización en hueso y dentina (Addison et al., 2010; Barros et al., 2013; Martin et al., 2008). Los miembros de la subfamilia SIBLING incluyen osteopontina (OPN), sialoproteína de unión a integrina (*IBSP*), proteína de matriz de dentina 1 (*DMP1*), sialofosfoproteína de dentina (DSPP) y fosfoglicoproteína de matriz extracelular (MEPE). Estas proteínas comparten ortólogos en tetrápodos pero eventos de translocación, pseudogénización o pérdida han sido particularmente relevantes y han modificado la posición de algunos ortólogos (Kawasaki & Amemiya, 2014). Tres genes pertenecientes al locus SCPP que codifican para proteínas SIBLING (*DMP1*, *IBSP* y *SCPPA2*) y que se expresan en hueso de *Xt* (Espinoza et al., 2010) son particularmente importantes para el presente trabajo, por lo que se hará una breve descripción de sus funciones y patrones de expresión.

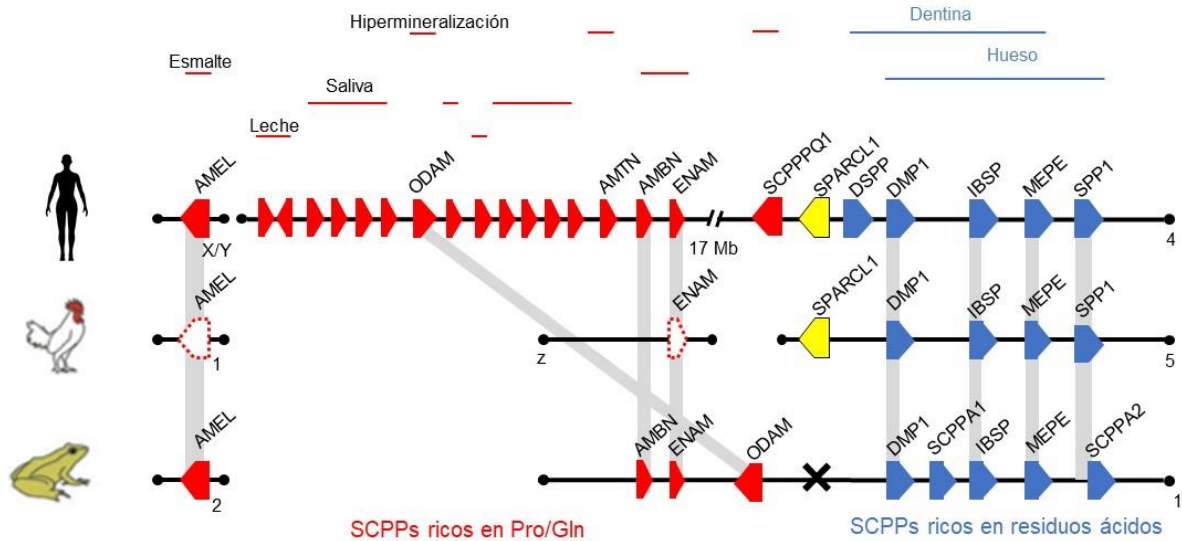


Figura 4: Organización genómica del locus SCPP

Arreglo y dirección transcripcional de los genes SCPP en los genomas de humano, pollo y rana muestran la organización en clúster. Genes ricos en Prolina (Pro) y Glutamina (Glu), ácidos y SPARCL1 son mostrados en rojo, azul y amarillo respectivamente. Los dominios de expresión y/o función son mostrados con líneas rojas y azules horizontales (arriba). Los ortólogos están unidos con líneas sombreadas. Las subsecuentes rondas de duplicación en tándem originaron el grupo de SCPP ricos en Pro/Gln en el humano. Observe la translocación cromosomal del gen ODAM en humano respecto a rana. SPARCL1 fue perdido en rana y eventos de pseudogenización en pollo son evidentes en los genes ricos en Pro/Glu (Imagen modificada de (Kawasaki & Amemiya, 2014)).

1.6.1 DMP1

La proteína *DMP1* es secretada durante la formación del hueso y diente. En mamíferos, su expresión ha sido identificada en osteoblastos que están al frente de la mineralización y en osteocitos (Kamiya & Takagi, 2001; Silvent et al., 2013). Su expresión temporal durante el desarrollo muestra que primero aparece en células de cartílago hipertrófico, osteoblastos (débilmente) y más tarde, fuertemente en osteocitos (Feng et al., 2002). La pérdida de función de esta proteína está asociada a raquitismo y osteomalacia. Esta última condición se refiere a un tejido pobremente mineralizado, lo que lleva también a una anormal regulación del fosfato derivado de una desregulación de los niveles circundantes del factor de crecimiento de fibroblastos (FG23) (Feng et al., 2006). Estas funciones llevan a la clasificación de *DMP1* como un *activador* en la vía regulatoria de la mineralización del esqueleto (Reznikov et al., 2020). Ratonés deficientes de *Dmp1* presentaron defectos en la mineralización en hueso y dentina (Ye et al., 2004) con características similares a las provocadas por mutaciones en humanos, los cuales desarrollan raquitismo hipofosfatémico (Maginot et al., 2015). Finalmente, la proteína

DMP1, ha sido observada en núcleos de preosteoblastos donde tiene un importante papel en la diferenciación. Esta función requiere de su transporte de la membrana plasmática al núcleo donde actuaría como un cofactor de la transcripción de genes responsables en la diferenciación de osteoblastos y odontoblastos (Merkel et al., 2019).

1.6.2 BSP

La sialoproteína de unión a integrina (*IBSP* o *BSP*) es una fosfoproteína específica de mineralización de hueso y diente. Estudios *in vitro* demostraron que actúa como un potente nucleador de la mineralización con capacidad de unión a colágeno tipo I. La proteína es caracterizada por ser una estructura flexible (Fisher et al., 2001; Vincent & Durrant, 2013), altamente glicosilada y sulfatada con dos dominios poliglutámicos que nuclean la hidroxiapatita en coordinación con varios residuos de serina fosforilados (Baht et al., 2008; Gordon et al., 2007). En hueso de mamíferos, *IBSP* está presente en osteoblastos maduros y en osteocitos. También actúa como un importante regulador tanto de la apoptosis así como la proliferación de condrocitos en la transición de cartílago al hueso durante el desarrollo de hueso endocondral (Holm et al., 2015). Ratones nulos para *IBSP* tienen retraso en el crecimiento y mineralización de huesos largos. Además, presentan una actividad reducida de recambio óseo como consecuencia de una afectación en la formación y actividad de osteoclastos (Malaval et al., 2008).

1.6.3 OPN

La osteopontina (*OPN*) es codificada por el gen *Spp1* en mamíferos. Dentro de las proteínas no-colagenosas, es una de las fosfoproteínas más abundantes y altamente modificada postraduccionalmente con alto contenido de aminoácidos cargados negativamente. En la vía regulatoria de la mineralización del esqueleto *OPN* es considerada una proteína inhibidora de la mineralización (Reznikov et al., 2020) por lo que se considera que guía el crecimiento del cristal, la forma, el tamaño y su alineamiento (Barros et al., 2013; Hunter, 2013; Pampena et al., 2004). La osteopontina fue detectada por inmunohistoquímica en ratas neonatas en la matriz extracelular de hueso y osteoide y celularmente en preosteoblastos, osteoblastos y osteocitos (Mark et al., 1987). Por hibridación *in situ*, se identificó también en condrocitos hipertróficos y odontoblastos (Nomura et al., 1988). En *Xenopus* no se ha detectado *Spp1* en el genoma, y el gen más próximo en cuanto a similitud y ubicación cromosomal es *Scppa2*, sin embargo, aunque éste no tiene similitud significativa de secuencia, su expresión tiene el mismo patrón que la *OPN*

de mamíferos (fuerte en osteoblastos y osteocitos) y puede adoptar funciones similares en la mineralización del hueso (Kawasaki, 2009).

1.6.4 SCPPA2

Aunque *DMP1*, *IBSP* y *SCPPA2* se expresan en odontoblastos presentan una clara diferencia de expresión en hueso de *Xt*. Los transcritos de *IBSP* se detectan en osteoblastos y débilmente en osteocitos, mientras que el gen *Dmp1* está apagado en osteoblastos y sufre una fuerte activación una vez que las células se diferencien a osteocitos, la última etapa del linaje osteogénico (Espinoza et al., 2010). *Scppa2* se expresa débilmente en odontoblastos y fuertemente en osteoblastos maduros y osteocitos, sugiriendo el uso de *Scppa2* principalmente en hueso y sólo débilmente en dentina (Kawasaki, 2009).

1.6.5 SPARC

Probablemente debido a su historia evolutiva común (Bertrand et al., 2013) SPARC presenta patrones de expresión similar a los SCPP (Espinoza et al., 2010; Kawasaki et al., 2007). Las proteínas de la familia SPARC preceden la emergencia de los vertebrados, unen calcio y actúan como chaperonas de colágeno. Entre los osteóctios la mayoría de las especies tienen dos parálogos *SPARC* y *SPARCL1* que son expresados en células esqueléticas (Bertrand et al., 2013). *SPARC* y *SPARCL1* se originaron en la primera ronda de duplicación del genoma completo. Ambos parálogos divergieron en vertebrados pero *SPARCL1* dio origen por duplicaciones en tándem a los SCPP en osteóctios. Una característica importante a destacar en este trabajo es que *SPARCL1* fue perdido en el genoma de *Xenopus* (Figura 4) (Kawasaki, 2009).

La proteína SPARC (Osteonectina), es principalmente expresada en tejidos mineralizados pero también se ha encontrado en una variedad de tejidos que muestran su naturaleza *semiubica* en vertebrados. Normalmente, su expresión está cercanamente relacionada con el colágeno tipo I. En hueso, se expresa fuerte en osteoblastos, y al igual que *IBSP* y el colágeno tipo I, sólo débilmente en osteocitos. SPARC contiene también sitios de unión a colágeno y una región de unión a hidroxapatita. Ratones nulos de SPARC se ven afectados en la formación de hueso, un evento relacionado al menor número de osteoblastos y osteoclastos, lo que lleva a una disminución de la remodelación con un balance óseo negativo que causa osteopenia (baja densidad de hueso) (Delany et al., 2000). Finalmente, la coemergencia de SPARC con colágeno en metazoos basales y la presencia de sitios de unión a

colágeno altamente conservados en SPARC sugieren que ambas proteínas tienen una relación funcional ancestral (Enault et al., 2018; Koehler et al., 2009; Morrissey et al., 2016).

1.7 Regulación transcripcional de los miembros SCPP

La osteogénesis en vertebrados involucra dos procesos de osificación: intramembranoso y endocondral. En el primero, células mesenquimales directamente diferencian en osteoblastos que dan lugar a la formación de hueso y en el segundo, la condensación de células mesenquimales genera un molde de cartílago, el cual posteriormente es reemplazado por tejido óseo mineralizado (Hall, 2005). Los factores de transcripción y vías de señalización involucrados en la diferenciación de células del linaje osteogénico son bien conocidos, y terminan activando genes que codifican para componentes clave de la matriz mineralizada tales como los colágenos fibrilares, SPARC, y los miembros SCPP (Long, 2011; Marcellini et al., 2012). Las células progenitoras (pre-osteoblastos), que dan origen a osteoblastos, expresan los factores de transcripción *Dlx5*, *Msx1* y *Runx2* (Komori, 2010; Otto et al., 1997; Stein et al., 2004). Posteriormente (río abajo de *Runx2*) se expresa el factor de transcripción *Osterix* (*Osx* también conocido como *Sp7*) el cuál es expresado en un estado temprano del linaje osteoblástico antes de que las células estén completamente diferenciadas e inicien la síntesis de proteínas de matriz tales como colágeno tipo I (Nakashima et al., 2002). La sucesiva activación de factores de transcripción en el proceso de diferenciación de células óseas es regulada por una compleja interacción de señales extracelulares (hormonas y ligandos) de diferentes vías de señalización que incluyen *Wnt*, *BMP*, *TGF β* , *FGF* y *Hh* (Hartmann, 2009; Marcellini et al., 2012).

Los genes especializados en secretar la matriz extracelular de hueso y que han sido mejor estudiados desde el campo de la regulación transcripcional son *Alpl* (fosfatasa alcalina), *Col1a1* y *Bglap* (osteocalcina) así como los miembros del locus SCPP: *Ibsp*, *Spp1* y *Dmp1*. De particular relevancia han sido los blancos transcripcionales asociados a *RUNX2*, ya que fueron identificados en regiones no codificantes flanqueando a los genes *Alpl*, *Ibsp* y *Spp1* (Stein et al., 2004). Sin embargo, varios estudios se han concentrado en regiones proximales de los promotores. Por ejemplo, secuencias consenso para *RUNX2* fueron analizadas a lo largo de un segmento a 8 kb río arriba del TSS de *Ibsp* en modelo murino. Dos de estas secuencias mostraron actividad en ensayos de transfección transiente (Paz et al., 2005). Ensayos de gen reportero con luciferasa revelaron una alta especificidad de expresión cuando se incluía en el vector un sitio de unión putativo a *Runx2* a 9 kb río arriba del TSS de *Ibsp* (Paz et al., 2005). Sin

embargo, otros estudios demostraron que esta misma región era inhibitoria (Zhang et al., 2009), lo que revela la complejidad de esta región promotora.

En los últimos años, estudios basados en secuenciación masiva tales como perfiles de expresión génica (RNA-seq) y de unión a sitios de transcripción o modificaciones de histonas por inmunoprecipitación de la cromatina (ChIP-seq) así como ensayos de cromatina accesible por transposasa (ATAC-seq) han permitido determinar una plétora de regiones regulatorias que muestran la naturaleza compleja de elementos reguladores en *cis* y *trans* en el genoma completo de células óseas. Así, se ha demostrado que el factor de transcripción *RUNX2* se une a promotores de genes de hueso y con mayor frecuencia a elementos reguladores a más de 2kb de distancia de los genes blanco, así como al extremo 3' de genes importantes en diferenciación osteogénica (Hâkelien et al., 2014). Específicamente, se identificaron regiones ricas de *RUNX2* en regiones distales (>10kb) flanqueando genes como *Spp1*, *Col1a1*, *Sp7*, *Alpl* y *Runx2* con mayor intensidad que en regiones proximales. Además, dichas regiones también fueron ricas en marcas epigenéticas importantes de enhancers putativos, por ejemplo, H3K4me1, H3K9ac y H3K9ac. Utilizando estas marcas de cromatina así como la unión de *RUNX2* y C/EBP β se identificaron 3 enhancers ubicados a -30, -20 y -10 kb río arriba del promotor del gen de la metaloproteinasas 13 (*Mmp13*) los cuáles fueron eliminados *in vitro* mediante CRISPR/Cas9. La eliminación del enhancer a -30 kb resultó en la completa supresión de *Mmp13* (Meyer et al., 2015). También se demostró la co-ocupancia del factor de transcripción C/EBP β con *RUNX2* así como la naturaleza dinámica de la unión de *RUNX2* en el proceso de diferenciación de osteoblastos (Meyer et al., 2014).

Respecto a los genes SCPP, un análisis de ChIP-seq en preosteoblastos y osteoblastos *in vitro* mostró que *Spp1* tiene sitios de unión a *RUNX2* así como marcas de cromatina asociadas a promotores activos en una ventana de 10 kb río arriba del TSS y un fuerte enriquecimiento en una región distal (~27 kb) que colocaliza con el FT C/EBP β así como con regiones ricas en marcas de histonas asociadas a enhancers activos tanto en preosteoblastos como en osteoblastos (Meyer et al., 2014). *Dmp1* tiene enriquecimiento de *RUNX2* a nivel del promotor basal con nula detección a distancias mayores a 2 kb, excepto cuando se evaluó la marca de cromatina H3K4me1 la cual, además de colocalizar con *RUNX2* en el promotor, se detectó también a ~8 kb río arriba del TSS. En otro estudio, *Ibsp* fue fuertemente enriquecido por *RUNX2* en el TSS y en una región a -12.5 kb (Tai et al., 2017).

1.8 Arquitectura regulatoria en osteoblastos

Pocos estudios se han publicado con énfasis en interacciones distales entre promotores y enhancers osteogénicos. Un análisis de enriquecimiento de RUNX2 en el genoma de osteoblastos *in vitro* mostró que este FT se une prioritariamente a regiones distales intergénicas e intrónicas (70%) y con menor frecuencia en TSS y promotores. Interesantemente, esta distribución fue similar a la proteína arquitectónica CTCF (Wu et al., 2014). Un estudio de hipersensibilidad de la ADNasa corroboró esta distribución en regiones intrónicas y en intergénicas distales a genes de hueso (Tai et al., 2017). Esto sugiere la posibilidad de la expresión de Runx2 en el contexto de la arquitectura tridimensional de la cromatina, posiblemente como sitios de nucleación para modificaciones estructurales indispensables para la expresión génica durante la osteoblastogénesis (Wu et al., 2014). Así, elementos reguladores distales pueden tener un papel regulatorio a través de interacciones a larga distancia probablemente mediante bucles donde la unión de Runx2 puede contribuir a la remodelación de la cromatina a través de su interacción con enzimas modificadoras de cromatina (Pelletier et al., 2002; Stein et al., 2004; Tai et al., 2017).

Sólo un trabajo específico en la estructura de la cromatina en osteogénesis ha sido publicado. En éste se utilizó la técnica de captura de cromosoma (3C) para identificar asociaciones a distancia entre las regiones promotoras de *Runx2* y *Supt3h* en varias líneas celulares de hueso llevando a la conclusión de una interacción dinámica y funcional durante la osteoblastogénesis. Los autores propusieron que el promotor *Supt3h* fue cooptado (o reclutado) como un enhancer por el promotor de *Runx2P1* (Barutcu et al., 2014). Esta asociación fue corroborada en el cromosoma 6 de humano (44198640-46071520) en una región >1MB que abarca ambos genes (*Runx2* y su bystander *Supt3h*) los cuales fueron localizados dentro de un GRB que predice la organización topológica del locus (Harmston et al., 2017).

Con todo lo anterior se demuestra que las diferentes técnicas de secuenciación masiva han permitido determinar regiones de cromatina distales que antes eran inaccesibles mediante el estudio de elementos únicamente proximales. Una ola de nuevos retos surge, por ejemplo, descubrir la funcionalidad de enhancers distales putativos y determinar sus genes blancos en el contexto de la diferenciación osteogénica así como los posibles mecanismos por los cuáles diferentes enhancers pueden reconocer a sus promotores específicos por ejemplo dentro de un RGB o en un TAD. Así, nuevos métodos de estudio de contactos entre regiones distales de la cromatina se han vuelto relevantes. Versiones modificadas del protocolo original de 3C (1 contra 1) han surgido (Dekker et al., 2002), tales como 4C, 5C (1 contra todos) y HiC (todos contra

todos) que permiten identificar regiones de interacción en el genoma. En este trabajo utilizamos el ensayo de captura de conformación de cromosoma circular (4C-seq) el cual es un método derivado de 3C designado para investigar la frecuencia de contactos en una región genómica de interés (Krijger et al., 2020; Stadhouders et al., 2013). Las tecnologías basadas en 3C dependen de un entrecruce o *crosslinking* de fragmentos de ADN que están espacialmente próximos en el núcleo, seguido de fragmentación del ADN usando enzimas de restricción y subsecuente ligación. Los fragmentos ligados generan moléculas híbridas que son sometidas a un proceso de “*de-crosslinking*” y purificadas para generar un templado 3C en el cuál los fragmentos ligados son detectados y cuantificados. En 4C-seq (Figura 5), el templado de 3C es digerido nuevamente con una segunda enzima de restricción seguido por una circularización de los fragmentos mediante un segundo paso de ligación (templado 4C). Posteriormente, se realiza una PCR inversa con partidores que hibridan en un fragmento de restricción seleccionado previamente (*viewpoint*) para amplificar los fragmentos ligados (*captures*). Dichos partidores son diseñados con adaptadores de secuenciación *Illumina* que permiten la generación y secuenciación de librerías (Krijger et al., 2020). De esta forma, se obtienen perfiles de contactos de alta resolución. La técnica ha sido empleada para identificar contactos enhancer-promotor en la arquitectura de la cromatina en desarrollo (Andrey et al., 2013; Dao et al., 2017; Noordermeer et al., 2011) y evolución (Acemel et al., 2016; Gehrke et al., 2015; Woltering et al., 2014), la formación de nuevos TADs y los principios espaciales del “secuestro” de enhancers en procesos patológicos (Gröschel et al., 2014; Lupiáñez et al., 2015) así como las regiones genómicas con propiedades de cromatina similares (De Wit et al., 2013; Wijchers et al., 2016).

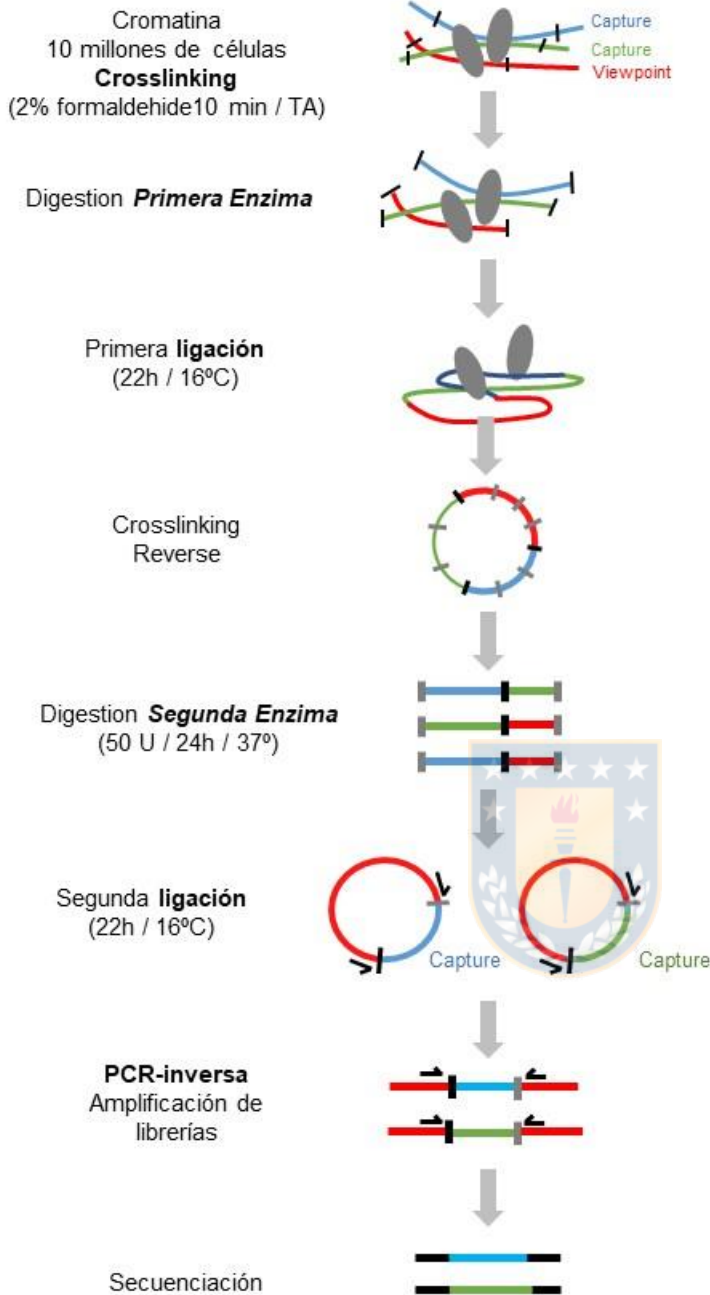


Figura 5: **Fundamento teórico del ensayo 4C-Seq**

Se utilizan 10 millones de células que se someten a un proceso de *crosslinking* con formaldehído por 10 min. La cromatina es digerida con una primera enzima de alta frecuencia de corte que genera extremos cohesivos, ligada y mediante *crosslinking* reverso liberada de proteínas (óvalos grises). Una segunda digestión con otra enzima de alta frecuencia de corte y subsecuente ligación en condiciones altamente diluidas permite generar moléculas circulares quiméricas. Este material se usa como templado para una PCR inversa que amplifica los fragmentos capturados, los cuáles son secuenciados y mapeados a un genoma de referencia. La razón del alto número celular es porque las topologías del genoma son variables entre células y entre alelos individuales de células únicas, por lo tanto, cada célula diploide sólo puede contribuir con un máximo de dos fragmentos de ADN ligados proximalmente (1 por cada alelo) para el perfil de contactos de un *viewpoint* dado. Por lo tanto, para amplificar las pares ligados, la PCR inversa necesita ser llevada a cabo en el equivalente de miles de genomas para generar perfiles de contactos que sean reproducibles y que expongan la mayoría de conformaciones dominantes en una población celular (Krijger et al., 2020).

2. PREGUNTA DE INVESTIGACIÓN E HIPÓTESIS

Los trabajos de regulación transcripcional de los genes SCPP realizados hasta la fecha se han limitado a estudiar regiones proximales de algunos miembros, como *IBSP* y *DMP1* en cultivos celulares *in vitro* y animales transgénicos (Balic & Mina, 2011; Gopalakrishnan et al., 2003; Higashibata et al., 2004; Paz et al., 2005; Sakuma et al., 2003). Sin embargo, tal como se detalló en la introducción, los genes involucrados en desarrollo y morfogénesis suelen tener enhancers que contactan a sus promotores a pesar de que están alejados en el genoma. **Identificar estos enhancers distales es fundamental para entender la arquitectura regulatoria del locus SCPP y permitirá explicar el hecho de que (1) genes organizados en tándem permanezcan agrupados por grandes periodos evolutivos y que (2) algunos genes del locus presenten patrones de expresión divergentes.** A diferencia de los estudios que se han enfocado principalmente en familias de genes que codifican para factores de transcripción, este proyecto apunta a entender la arquitectura regulatoria de locus SCPP cuyos miembros codifican componentes clave de matriz del tejido esquelético mineralizado en vertebrados.

El conocimiento actual de la transición osteoblasto-osteocito es escaso: a la fecha se desconoce si (1) *IBSP* y *SCPPA2* están siendo expresados activamente en osteoblastos y en osteocitos, o si (2) están siendo expresados activamente únicamente en osteoblastos, pero por la vida media larga de los transcritos, se siguen detectando en osteocitos. En la extremidad posterior del embrión de pollo, por ejemplo, los ARN mensajeros de *Fgf8* permanecen presentes en células de los somitas mucho después de que el gen se haya apagado transcripcionalmente, y sus niveles van decayendo a medida que las células se alejan del brote de cola ubicado en el extremo posterior (Dubrulle & Pourquié, 2004). Esta situación es similar a lo observado con los transcritos del gen *Col1a1*, que se expresa fuertemente en osteoblastos, y cuyos niveles disminuyen a medida que los osteocitos se encuentran embebidos más profundamente en la matriz (Enault et al., 2015). Esta similitud con *Fgf8* sugiere que el gen *Colágeno 1a1* se apaga en la transición osteoblasto/osteocito, algo que no ha sido comprobado. Por lo tanto, **determinar en qué tipo celular exactamente se transcriben los genes *IBSP* y *SCPPA2* es fundamental para entender si los MCRs del locus SCPP están activos sólo en osteoblastos, o tanto en osteoblastos como en osteocitos.** En este trabajo se incluirá también el gen *SPARC*, como paradigma para la función ancestral del locus SCPP.

Con base en los antecedentes detallados en la introducción planteamos como pregunta de investigación general: **¿Se caracterizará el locus SCPP por la presencia de múltiples**

MCRs que contactan y corregulan genes coexpresados en la diferenciación del linaje osteogénico? Esta pregunta ha originado la siguiente hipótesis: **la activación transcripcional del locus SCPP involucra la acción de múltiples MCRs que contactan y coregulan promotores de miembros coexpresados en etapas específicas del proceso de diferenciación del linaje osteogénico.**



3. OBJETIVO GENERAL

Determinar si la transcripción de genes del locus SCPP involucra la acción de múltiples MCRs que contactan y corregulan miembros coexpresados en etapas específicas del proceso de diferenciación del linaje osteogénico en *Xenopus tropicalis*.

4. OBJETIVOS ESPECÍFICOS

1. Determinar qué tipos celulares transcriben activamente los genes SCPP acídicos (*DMP1*, *IBSP* y *SPPA2*) y *SPARC* durante la osteogénesis.
2. Caracterizar la actividad de los MCRs putativos de los genes SCPP y *SPARC* en osteoblastos.
3. Identificar los MCRs que contactan a los promotores basales de los genes SCPP y *SPARC* en osteoblastos.



5. MATERIALES Y MÉTODOS

5.1 Animales de experimentación y preparación de muestras

Todos los experimentos fueron realizados usando material biológico del anfibio *Xt* para lo cual se cuenta con un bioterio en la Universidad de Concepción. Los organismos adultos fueron mantenidos de acuerdo a los protocolos estándar para la especie. Se utilizaron renacuajos en el estadio Nieuwkoop and Faber (NF) 58-60 tanto para la realización de cultivos primarios de células óseas, así como para obtener tejidos para histología e hibridaciones *in situ*. Los embriones se obtuvieron mediante procesos de fertilización inducida por 200U de la hormona Gonadotrofina Coriónica Humana (HCG) (Intervet International B.V) y se mantuvieron hasta alcanzar el estadio requerido para los experimentos. Para la obtención del material biológico los renacuajos fueron anestesiados con triclaína 200mg/ml (Sigma–Aldrich, St. Louis, MO, USA) y disectados en medio HSBB 0.6x (Invitrogen, Paisley, UK) ya sea para la obtención de células óseas de calvaria (hueso frontoparietal de renacuajo) o para la extracción de tejidos de extremidades posteriores que fueron inmediatamente fijadas con PFA al 4%. Los procedimientos experimentales para este trabajo fueron aprobados por el Comité de Bioética de la Universidad de Concepción y efectuados siguiendo las pautas establecidas en el manual de bioseguridad y bioética de la Agencia Nacional de Investigación y Desarrollo, ANID de Chile.

5.2 Análisis de expresión espacial por hibridación *in situ*

5.2.1 Amplificación de regiones intrónicas

Las sondas intrónicas (~600 pb) de los genes *Dmp1*, *Ibsp*, *Scppa2* y *Sparc* fueron amplificadas desde ADN genómico. Los productos de PCR fueron ligados a vectores pBluescript (pBS), los cuáles fueron linearizados a través de reacciones de restricción, utilizando 10 UI de enzima y 10 µg de ADN en 50 µl de reacción, e incubados a 37°C por toda la noche. Una vez linearizados se agregaron 50µl de TE (pH 8.0), 5µl de SDS 10% y 0.8µl de PK 10mg/ml y se incubaron a 37°C. Posteriormente, los plásmidos fueron precipitados, para ello se agregó un volumen de fenol/cloroformo 1:1, se mezcló vigorosamente por 3 minutos, y se centrifugó a 14.000rpm por 5 minutos a 4°C. Se extrajo la fase acuosa, y se transfirió a un nuevo tubo y se repitió la extracción fenol/cloroformo una vez más. Después, se agregó un volumen de cloroformo 1:1, y se repitió el proceso de centrifugación. Finalmente se añadieron 1/10 de volumen de acetato de sodio 3M pH 5.5 y el doble de volumen de etanol 100%. Todo fue mezclado por inversión y precipitado a -

20°C por 30 minutos. Una vez precipitado, el ADN fue centrifugado a 14.000rpm por 10 minutos a 4°C, el sobrenadante fue removido y el pellet fue lavado suavemente con 1ml de etanol 70%. La muestra fue nuevamente centrifugada a 14.000rpm por 3 minutos a 4°C. El sobrenadante fue removido, y los tubos secados a 37°C por 5 a 10 minutos. Finalmente, el pellet de ADN fue resuspendido en 10µl de Agua-DEPC, y guardado a -20°C para su posterior uso en reacciones de transcripción.

5.2.2 Reacciones de transcripción *in vitro*

Los plásmidos pBS linearizados que contenían las regiones genómicas intrónicas de los genes SCPP y SPARC fueron utilizados como molde para sintetizar sondas Dig-ARN sentido y antisentido (Roche). La reacción de transcripción fue ensamblada en agua libre de RNAsas en un volumen de 10µL. Se utilizaron 1µL de polimerasa de ARN T3 o T7, 1 µL de plásmido linearizado (~1ug), 1µL of buffer de transcripción 10X, 1µL of DIG Labeling Mix y 0.5 µL of RNAsine (RNase inhibitor Promega). La reacción fue incubada 2h a 37°C. Posteriormente se añadieron 6µL of agua DEPC, 2 µL of buffer de DNase RQ1 y 2µL de DNase RQ1 (Sigma). La reacción se incubó 30 min a 37°C. Para la purificación de sondas se agregó 1µL of EDTA 0,5 M and 2,5 µL de LiCl 4M. Se mezcló por inversión y se agregaron 75µL de ethanol frío al 100%. La reacción se incubó a -20°C toda la noche. Al día siguiente se centrifugó a 4°C, 13000 rpm durante 20 min. Se reemplazó el sobrenadante por etanol frío al 70% y se centrifugó a 4°C durante 10 min. El pellet fue secado a TA y se resuspendió en 10 µL de agua DEPC. 1 µL de reacción fue cargado en un gel de agarosa.

5.2.3 Hibridación in situ con sondas intrónicas y exónicas

Sondas intrónicas y exónicas de los genes SCPP (*Dmp1*, *Ibsp* y *Sppa2*) y *Sparc* fueron utilizadas para la realización de hibridaciones in situ en muestras de fémur de renacuajos en el estadio NF60 que fueron descalcificadas, incluidas en parafina y seccionadas en 7 y 10 µm. Los cortes fueron desparafinados, deshidratados progresivamente y tratados con Proteínasa K (10 µg/ml) a 37°C por 30 minutos para facilitar el ingreso de las sondas ARN-Dig. Subsiguientemente, las secciones fueron tratadas con trietanolamina (0.1 M, pH 8) y anhídrido acético 0.25%. Después las muestras fueron incubadas en solución de hibridación a 60°C toda la noche. Al día siguiente se realizaron lavados y un tratamiento con RNasa A para eliminar los excesos de sondas ARN-Dig sin hibridar. Las muestras fueron lavadas, bloqueadas e incubadas con anti-DIG-AP 1:2000

toda la noche dentro de una cámara húmeda. Al tercer día las muestras fueron lavadas y reveladas con una solución colorante saturada de NBT/BCIP e incubadas a temperatura ambiente para las sondas exónicas y a 4°C para las intrónicas. El proceso de revelado fue terminado cuando la señal fue visualizada. Todas las hibridaciones incluyeron una sonda sentido intrónica y una exónica como controles negativos. Finalmente, las muestras fueron fijadas y deshidratadas progresivamente para ser teñidas con DAPI y montadas con solución entellan. Todas las muestras fueron observadas en un microscopio de campo claro. Imágenes representativas fueron obtenidas por microscopía de campo claro y epifluorescencia.

Las secuencias de los partidores para el diseño de las ondas intrónicas de todos los genes de estudio, así como para la sonda exónica del gen *Scppa2* se encuentran en el anexo 1.

5.3 Selección de MCRs putativos

Para determinar la expresión de los genes SCPP así como los MCRs putativos del locus SCPP y del gen SPARC se recurrió a datos obtenidos por estudios de RNA-seq y CHIP-seq realizados en nuestro laboratorio (Hanna, 2014). Se identificaron los genes SCPP que se expresan en hueso así como el enriquecimiento de las marcas de histonas H3K4me1 y H3K27ac para determinar enhancers putativos en una ventana de 336 Kb. Las marcas H3K27ac y H3K4me3 y la señal por RNA-seq del primer exon, fueron importantes para reconocer los promotores cercanos a los genes blanco *Ambn*, *Dmp1*, *Ibsp*, *Scppa2*. El mismo procedimiento fue realizado para el gen SPARC que abarca una ventana de 150 Kb (gen más marcas de cromatina en regiones intergénicas). Se utilizó la base de datos gbrowse (<http://localhost:4321/cgi-bin/gbrowse/gbrowse/EXP2/>) para visualizar los perfiles de expresión y marcas de CHIP-seq de los genes de interés en el genoma de referencia de *Xt* versión 4.1 (<http://www.xenbase.org>).

Se seleccionaron 4 enhancers putativos de mayor enriquecimiento río arriba del promotor basal de cada uno de los genes SCPP (*Dmp1*, *Ibsp*, *Scppa2*). Dichos enhancers fueron identificados mediante su posición relativa en el genoma con los números 244, 280, 345 y 364. En la ventana analizada para el gen SPARC se identificaron 7 enhancers putativos (100, 151, 159, 175, 189 y 202). Para los controles negativos se seleccionó una región del genoma con nulo enriquecimiento de marca de histonas, conservación o repeticiones. Esta región fue elegida de una intrón del gen AMBN y se denominó control negativo intrónico (CNI).

Después de elegir los enhancers con mejor ranking (mayores niveles de expresión relativa) próximos a los genes SCPP y SPARC se procedió a analizarlos in silico para el diseño de partidores. Se identificaron los exones e intrónes de cada gen de interés y se determinó el

sitio de inicio de la transcripción. Se descargaron las secuencias de los TSS (~0.5 kb) y de los MCRs putativos seleccionados (~1-2 kb). Se realizó un análisis de las repeticiones tanto internas del fragmento a clonar (en tándem) como las genómicas. Mediante la herramienta *autoblast* identificamos las repeticiones internas y con Blast/Blat buscamos las repeticiones en todo el genoma. Posteriormente se evaluó la conservación del promotor comparando *Xt* con el genoma de *Xenopus leavis* con la herramienta de bioinformática VISTA.

5.4 Generación de plásmidos para ensayos reporteros

5.4.1 Amplificación y purificación de promotores basales y enhancers

Se amplificaron los promotores basales (TSS) de los genes SCPP (*Dmp1*, *Ibsp*, *Scppa2*) y del gen *Sparc*, así como todos los enhancers putativos del locus SCPP (244, 280, 345 y 364) y del gen *Sparc* (100, 151, 159, 175, 189 y 202) a partir de ADN genómico extraído de larvas de *Xt*. La amplificación fue realizada utilizando la enzima polimerasa KOD (Calbiochem, San Diego, CA, USA), siguiendo las instrucciones del fabricante. Los amplicones fueron purificados mediante el kit Wizard SV Gel and PCR clean-up System (Promega, Madison, USA). El ADN purificado fue cuantificado mediante Nanodrop y almacenado a -80°C.

5.4.2 Ligación, transformación y purificación de plásmidos CisTA

Las regiones de los promotores basales (TSS) fueron ligadas en un vector de clonación Cis-TA mediante la enzima ADN ligasa (T4 DNA Ligase, 5 Weiss U/μl, Thermo Scientific) siguiendo las instrucciones del fabricante. En breve, los productos de PCR de los promotores basales de cada gen y el vector CisTA linearizado previamente con la enzima *HincII* fueron sustratos para la reacción de ligación (Figura 6A). La reacción se llevó a cabo con 30 ng de vector CisTA y 90ng de inserto (1:3) y fue incubada 2 hrs a 22°C, posteriormente se inactivó incubando a 65°C por 10 min. Los plásmidos generados (pCisTA-TSS + nombre de cada gen) fueron ingresados en bacterias quimio-competentes DH5α por la técnica de transformación bacteriana utilizando 3.5 μl de reacción en 50 μl de bacterias, incubadas a 42°C durante 35s e incubadas en hielo por dos min. Posteriormente se recuperaron con 350 μL de medio LB temperado a 37°C. Las bacterias transformadas fueron incubadas por 1h en agitación a 37°C, sembradas concentradas (100 μl) y diluidas (1:10) en placas con Xgal 1mg/μl preparadas previamente con antibiótico ampicilina como medio de selección. Las placas se incubaron toda la noche a 37°C.

Al día siguiente se seleccionaron las colonias blancas (que ingresaron el inserto) y se evaluaron por PCR colonia para determinar la dirección sentido con los partidores forward del inserto y M13-reverse del vector CisTA. Los clones correctamente orientados se inocularon en medio LB con ampicilina y se incubaron a 37°C con agitación (200 rpm) toda la noche. Al día siguiente se realizó miniprep kit (Promega) para purificar cada plásmido a partir de 4 ml de bacterias. Los plásmidos fueron cuantificados y analizados con diferentes enzimas de restricción para corroborar su naturaleza nucleotídica. Una vez que los plásmidos fueron confirmados correctamente, se almacenaron a -20°C para su posterior uso en reacciones de recombinación Gateway BP.

5.4.3 Generación de vectores ENTRY

Las regiones de ingreso de cada promotor basal en el vector Cis-TA están flanqueadas por secuencias attB3 y attB5 diseñadas en nuestro laboratorio para reacciones de recombinación Gateway. Dicho sistema de clonamiento permite ingresar fragmentos de interés en vectores de entrada (Entry) mediante la enzima clonasa II BP. En esta reacción se recombinan *in vitro* los sitios attB del vector CisTA y un vector donante (pDonor201) p3-p5 que contiene un casete génico del gen suicida ccdB y otro de resistencia a cloranfenicol CmR. Estos casetes están flanqueados por los sitios de recombinación attP3 y attP5, y el backbone de este plásmido contiene el gen de resistencia a Kanamicina (Figura 6B).

Una vez ingresado el promotor basal de cada gen (~500 pb) en un vector Entry, es posible linearizar el plásmido Entry en un sitio río arriba del TSS (Por ejemplo, con la enzima *SmaI*) e ingresar cada uno de los diferentes enhancers. De tal forma que cada uno de los promotores basales está listo para ser combinado con cualquier enhancer de interés, lo que le da versatilidad, y eficiencia a este sistema de clonamiento (Figura 6C). Dichos vectores (pEntry-Enhancer:TSS) son posteriormente recombinados con un plásmido destino funcional (Destino) mediante la enzima clonasa LR lo que permite generar plásmidos de trasfección (Figura 6D). El vector destino ha sido modificado previamente para contener un control interno de transfección cuyo promotor es CMV acoplado al gen reportero mCherry fusionado a histona y una caseta donde se ingresa cada promotor basal con su enhancer acoplados al marco de lectura de la proteína verde fluorescente GFP nuclear (Figura 6D).

5.4.4 Reacción BP

Los plásmidos Donor p3-p5 y cada vector TSS-CisTA fueron sustratos para la reacción BP, la cual fue incubada a 25°C toda la noche. El plásmido generado (pEntry-TSS-Nombre de cada gen) fue ingresado en bacterias quimio-competentes DH5α por transformación (42°C durante 45s) en una relación de 3.5 μL de reacción en 50 μL de bacterias. Después de 30 min en hielo, las bacterias se recuperaron con 450 de medio LB y se centrifugaron a 1200 RPM. Se eliminaron 350 μL de sobrenadante y el resto (100 μL) fue sembrado en placas que contenían el antibiótico kanamicina. Las placas se incubaron durante 16 hrs a 37°C. Al día siguiente todas las colonias fueron evaluadas por PCR para determinar el ingreso del inserto utilizando un partidior forward del vector Entry y un partidior reverse del inserto. Los clones positivos se inocularon en 5 ml de medio LB con kanamicina y se incubaron a 37°C con agitación (200 rpm) toda la noche. Al día siguiente se realizó miniprep con un kit (Promega) para purificar los plásmidos a partir de 4 ml de bacterias. Los plásmidos fueron cuantificados y analizados con diferentes enzimas de restricción para corroborar su naturaleza nucleotídica. Una vez que los plásmidos fueron confirmados correctamente, se almacenaron a -20°C para su posterior uso en reacciones de recombinación Gateway LR.

5.4.5 Ligación de enhancers en el vector Entry

Los productos PCR de cada enhancer y de un intrón del gen AMBN (control negativo intrónico; CNI) así como cada plásmido Entry que contiene el TSS de cada gen fueron sustratos de las reacciones de ligación. Cada reacción se llevó a cabo en una relación 1:3 (vector-inserto) utilizando 30 ng de vector cisTA y 90 ng de inserto. La reacción fue incubada 2 hrs a 22°C, posteriormente se inactivó incubando a 65° por 10 min. Los plásmidos Entry generados (Enhancer-TSS de cada gen) fue ingresado en bacterias quimio-competentes DH5α por transformación (42°C durante 45s) en una relación de 3.5 μL de reacción en 50 μL de bacterias (recuperación con 350 μL de medio LB) y estas fueron plaqueadas (100 μL) y en una dilución 1:10 y seleccionadas mediante el antibiótico kanamicina. Las placas se dejaron incubando a 37°C toda la noche. Al día siguiente se evaluaron los clones mediante PCR colonia, los clones positivos fueron inoculados y procesados como se describió en el apartado anterior.

5.4.6 Reacción LR

Los plásmidos Entry-TSS o las diferentes combinaciones Entry-Enhancer:TSS así como el vector Destino fueron sustratos para la reacción LR, para la cuál se utilizaron 75 ng del vector Entry y 150 ng del vector Destino; 2 μ L de enzima clonasa LR y 2 μ L de buffer 5X en una reacción de 10 μ L con buffer TE. La reacción fue incubada a 25°C durante 2 hrs. El plásmido generado CMV:mCherry-Enhancer-TSS-nombre de cada gen:GFPn fue ingresado en bacterias quimio-competentes DH5 α por transformación (42°C durante 30s) en una relación de 3.5 μ L de reacción en 50 μ L de bacterias y estas fueron plaqueadas y seleccionadas mediante el antibiótico ampicilina. Las placas se dejaron incubando a 37°C O/N y al día siguiente los clones fueron comprobados mediante PCR colonia utilizando un partidor forward del inserto, así como un partidor reverse que se encuentra en el marco de lectura de GFP. Los clones positivos fueron inoculados en 5 ml de medio LB con ampicilina y se incubaron a 37°C con agitación (200rpm) durante 1 h. Al día siguiente se realizó miniprep kit (Qiagen) para purificar ambos plásmidos a partir de 4 ml de bacterias y se realizaron los respectivos ensayos de restricción para comprobar la naturaleza nucleotídica de los plásmidos de expresión. Todos los plásmidos comprobados fueron anotados y almacenados a -20°C.



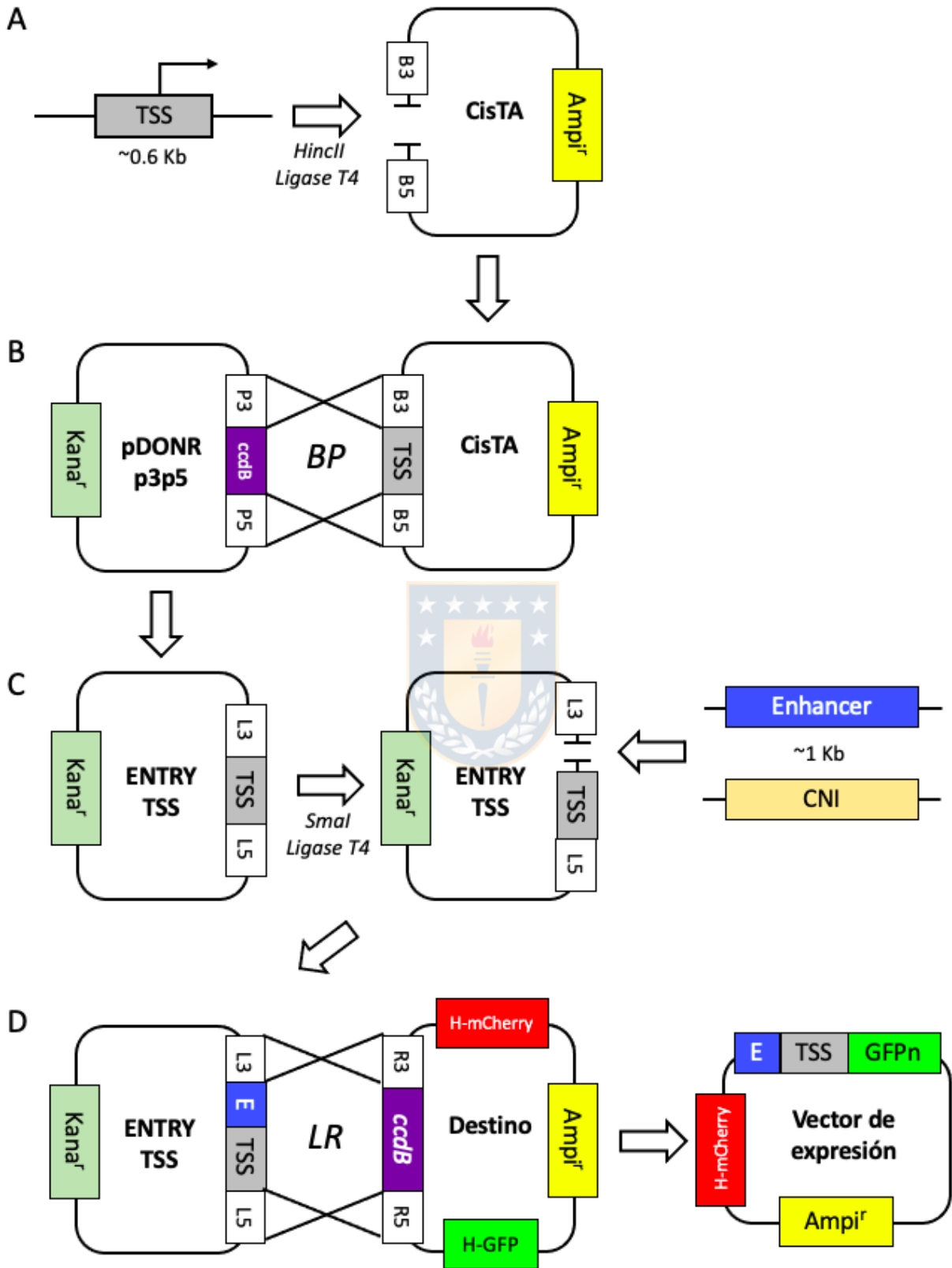


Figura 6: Generación de vectores de expresión mediante el sistema Gateway

En A, los promotores de los genes de estudio fueron ingresados a un vector CisTA previamente linearizado con la enzima *HincII*. Dicho vector cuenta con los sitios de recombinación attB3 y attB5 que al recombinarse con un vector Donnor mediante una reacción con la clonasa BP permite generar un vector Entry que es resistente al antibiótico Kanamicina. Este último es linearizado en un sitio *SmaI* al cuál es ligado cada uno de los enhancers de estudio, incluyendo una región control CNI (Control Negativo Intrónico). La ligación permite ingresar los diferentes MCRs río arriba de cada promotor basal. Una subsiguiente reacción con la clonasa LR permite intercambiar el caset del gen suicida (ccdB) con el fragmento de interés que contiene el enhancer a estudiar acoplado a su promotor basal, los cuáles son ingresados finalmente a un plásmido de expresión que contiene mCherry como control interno y el marco de lectura de la GFP dirigido por los elementos *cis* reguladores de interés.

5.5 Ensayos funcionales con reporteros

Por alguna razón fisiológica asociada a las características de los osteoblastos, no fue posible alcanzar una adecuada eficiencia utilizando diferentes métodos de transfección (*p. ej.* Lipofectamina (Invitrogen), Fugene HD (Promega) y Attractene(Qiagen)) por lo que se procedió a utilizar el método de transfección con el sistema de electroporación Neon propuestos por (Bertin et al., 2015). Desafortunadamente, las condiciones estandarizadas hasta ese momento utilizaban una gran cantidad celular (1 millón por condición) con una eficiencia de hasta 15% pero con muy baja viabilidad (menor al 50%). Esto generaba dos problemas: el uso de una gran cantidad de osteoblastos primarios era prácticamente imposible de obtener si se querían probar diferentes plásmidos y la baja viabilidad celular que imposibilitaba la reproducibilidad de los experimentos, además de las grandes cantidades de DNA de cada plásmido que tenían que purificarse para cada condición. Por lo tanto, un enorme reto era estandarizar la electroporación utilizando la menor cantidad celular sin comprometer la eficiencia.

5.5.1 Estandarización de electroporación mediante el sistema Neon

Se empezó realizando una curva de calibración cercana a los datos propuestos por (Bertin et al., 2015) combinando las diferentes condiciones de 1) Voltaje, 2) Tiempo (ms), No. pulsos (1,2,3), Cantidad de ADN (ng) y una densidad celular de 70,000 células con el objetivo de alcanzar una eficiencia de al menos entre 10-15% para detectar y cuantificar la actividad de los diferentes promotores de estudio. Para la estandarización se transfectó el plásmido CMV-mCherry. La fluorescencia de célula única fue cuantificada mediante citometría de flujo 48 hrs después de electroporar. Se utilizó el canal PE-Texas red-A para detectar fluorescencia de la proteína mCherry. Todas las muestras fueron analizadas a una velocidad lenta detectando un

total de 1000 eventos. Los datos de mejor eficiencia propuestos por (Bertín et al., 2015) V=1230V, T=30 ms, 2 pulsos fueron corroborados y se posicionaron entre las tres mejores condiciones de electroporación (~14%). Un incremento en el número de pulsos mejoró la eficiencia hasta un 20.7%, sin embargo, la viabilidad continuó siendo muy baja lo que llevó algunas veces a cuantificar menos de 1000 eventos por citometría, nuevamente, dificultando la reproducibilidad y el análisis estadístico (Ver tabla 1). Finalmente, modificar estas condiciones en su cercanía llevó inevitablemente a la inviabilidad celular o a la disminución drástica de la eficiencia.

	VOLTAJE	TIEMPO (ms)	Pulsos	[DNA] ug	No Cel
A1	9,1 %	1120	30	2	1 70000
A2	13,6 %	1170	30	2	1 70000
A3	15,4 %	1220	30	2	1 70000
A4	11,9 %	1120	30	3	1 70000
A5	15,3 %	1170	30	3	1 70000
A6	20,7 %	1220	30	3	1 70000

Table 1: Optimización de la electroporación

La condición A6 mostró las mejores condiciones de electroporación de osteoblastos con altas eficiencia (20,7 %) y viabilidad. La condición A3 era usada para electroporar 1 millón de células y también se posicionó con una de las mejores eficiencias. Los resultados resumidos en esta tabla fueron el resultado de la electroporación en todas las condiciones del plásmido CMV:mCherry. Todos los datos fueron obtenidos a partir de 70,000 células sembradas en una caja p24 y 1000 eventos cuantificados 48 h después de la electroporación. Todos los datos fueron obtenidos por duplicado.

Si el aumento en el tiempo y el número de pulsos era una condición inviable (hasta un 50%; datos no mostrados), la siguiente condición lógica sería alterar el voltaje, por lo cuál se procedió a una nueva curva de calibración completa con un voltaje mínimo posible de 850 y un máximo de 1700 alterando la amplitud y los pulsos de acuerdo a las condiciones propuestas por el fabricante y una cuantificación precisa de la viabilidad (Figura 7).

Sorpresivamente, las condiciones 5, 9 y 12, que son totalmente diferentes a las estandarizadas previamente, mostraron ser las mejores en cuanto a eficiencia y viabilidad (Figura 7). Además, todas las condiciones cercanas a lo propuesto para 1 millón de células (25-30), eran inviables para 70,000 células.

Con este proceso de estandarización se encontró la mejor condición (número 5) que nos llevó a la mejor eficiencia (38%) y viabilidad (85%). Así logramos electroporar sólo 70,000 células utilizando sólo 1 µg de ADN por condición y cuantificando hasta 5000 eventos por citometría. Estas son las condiciones que actualmente se utilizan en nuestro laboratorio para electroporar

cualquier plásmido, lo que ha llevado a tener éxito en experimentos de ensayos con reporteros. Es importante mencionar que eficiencias h

asta del 40% son obtenidas con plásmidos pequeños (CMV:mCherry por ejemplo), sin embargo, plásmidos de mayor tamaño que contienen la caseta CMV-mCherry y un promotor acoplado a su enhancer dirigiendo la expresión de GFP son mucho menos eficientes. No obstante, adecuados aún para cuantificar ensayos con reporteros.

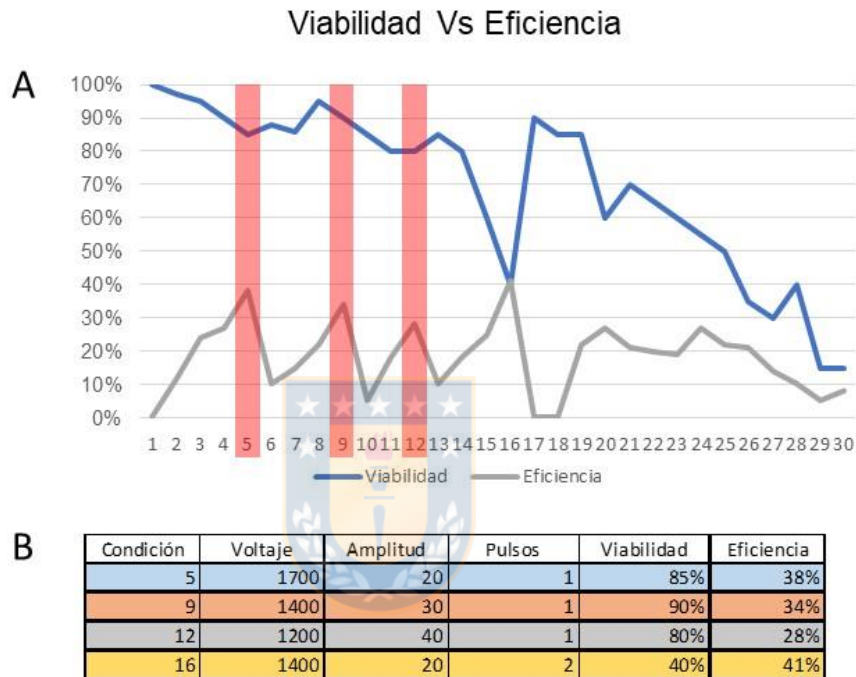


Figura 7: Calibración de las condiciones de electroporación

En A, se grafican las diferentes condiciones probadas en el proceso de estandarización de la electroporación utilizando 70,000 células y 1 µg de ADN por condición. En el eje de las Y se muestra el porcentaje de eficiencia (línea gris) contra viabilidad (azul). Las condiciones 5, 9, 12 y 16 mostraron las mejores eficiencias, no obstante, la condición 16 disminuyó la viabilidad drásticamente, probablemente por el aumento de un pulso al momento de electroporar. En B, el resultado resumido de probar 30 diferentes condiciones con los resultados de las 4 mejores. La eficiencia fue cuantificada mediante la fluorescencia única de células rojas (mCherry) en el cuadrante de células rojas y la viabilidad cuantificando el número de células vivas después de electroporar divididas por el total de vivas y muertas (azul de tripano) y multiplicando el resultado por 100. La estandarización se realizó por duplicado con dos réplicas experimentales.

5.5.2 Cultivo primario de osteoblastos

El cultivo de células primarias de osteoblastos fue realizado de acuerdo al protocolo propuesto por (Bertin et al., 2015). Las células fueron extraídas del hueso frontoparietal (calvarias) de renacuajos de *Xt* en el estadio NF58-60. Para cada cultivo se utilizaron al menos 20 calvarias (~20,000 células por calvaria) que fueron disectadas y recolectadas en 2 tubos *ependorf* de 2ml con 500 μ l de medio HBSS 0.6x (Gybco) con 1% Penicilina/Estreptomicina (P/E) (Hyclone). Con la finalidad de eliminar células pigmentadas, cristales de hidroxiapatita, eritrocitos u otros tipos celulares de tejido blando las calvarias fueron tratadas con 200 μ l de colagenasa P al 0.1% en HBSS 0.6x 1% P/S por 40 minutos a TA con agitación suave. Después de eliminar el sobrenadante se procedió a realizar un tratamiento con 200 μ l de tripsina 0.125% y colagenasa P 0.1% en HBSS 0.6x 1% P/S por 10 minutos a TA con agitación suave. Posteriormente, se llevaron a cabo tres tratamientos de tripsina 0.25%, y colagenasa P 0.1% en HBSS 0.6x 1% P/S por 15 minutos a TA con agitación suave. En cada tratamiento, los osteoblastos fueron recolectados en 5ml de medio L15 0.6X (Gibco) con 10% SBF, 1% P/S, 1% L-Glu, 25 μ g/ml de Gentamicina y 0.25 μ g/ml Fungizona en un tubo falcon de 15ml para inactivar la tripsina. Las células fueron centrifugadas a 1100rpm por 5 minutos, resuspendidas en 2ml de medio L15 0.6x suplementado con 10% SBF, 1% L-Glu y 1%P/E y cuantificadas utilizando una cámara de Neubauer. Un total de 400,000 células fueron sembradas en placas de 35mm e incubadas a 28°C. Para una correcta asepsia, se realizó un cambio de medio 4 hrs después del sembrado. Al ser un tipo celular con fenotipo pseudoepitelial es posible que los osteoblastos lleguen a confluencia 72 hrs después de la extracción para su posible uso o expansión.

5.5.3 Electroporación de reporteros del locus SCPP y SPARC

Los plásmidos de expresión construidos en el apartado 4 de este trabajo fueron electroporados utilizando la condición 5 (1700V - 20ms - 1pulso) propuesta en el apartado de estandarización de la electroporación. Antes de iniciar el experimento se prepararon alícuotas de todos los plásmidos a electroporar a una concentración de 600 ng/ μ L, se temperó el medio de crecimiento sin antibiótico, tripsina EDTA, PBS 0.6X, buffer B o T. Se preparó también el equipo Neon y las puntas de electroporación estériles de 10 μ L. 2 μ L de plásmido (1.2 μ g) fueron colocados en tubos *ependorf* de 0.6 ml. Posteriormente, células confluentes expandidas en placas p100 (preferentemente de pasaje 1 o 2) fueron lavadas dos veces con PBS 0.6X, tripsinizadas, cuantificadas y recuperadas con 5 ml de medio sin antibiótico. El total de células fue dividido

entre el total de posibles electroporaciones. Tomando en cuenta que una placa p100 puede contribuir con 2 millones de células, entonces es posible electroporar hasta 20 plásmidos, sin embargo, no es recomendable, ya que la viabilidad es afectada negativamente con el tiempo que las células permanecen en buffer B. Por lo tanto, por cada placa p100 sólo fue posible realizar un máximo de 10 electroporaciones. Así, se cuantificaron 1 millón de células que fueron centrifugadas y resuspendidas en 100 μ L de buffer B. El electroporador fue programado en la condición 5 previamente estandarizada. Se utilizaron 10 μ L de células resuspendidas (aprox. 70,000 células debido a la pérdida por centrifugación) con pipeta Neon y se mezclaron en el tubo previamente preparado con 2 μ L de plásmido (concentración final de 1 μ g/ μ L) y se procedió a electroporar. Inmediatamente las células se resuspendieron en 500 μ L de medio completo sin antibiótico previamente temperado y depositado en tubos *ependorf* de 500 μ L. Todas las electroporaciones se realizaron por duplicado y siempre con su respectivo control (células electroporadas sin plásmido). Después de recuperadas las células fueron removidas de cada tubo *ependorf* para ser sembradas en placas de 24 pocillos. 4 hrs después el medio fue cuidadosamente removido e intercambiado muy suavemente por medio con antibiótico previamente temperado. Las células fueron incubadas a 28°C por 48 h.

5.5.4 Cuantificación de fluorescencia por citometría de flujo

Los datos de fluorescencia de células únicas fueron cuantificados en el citómetro de flujo Fortessa X20, equipo facilitado por el laboratorio de inmunovirología a cargo de la Dra. María Inés Barría de la facultad de Ciencias Biológicas de la Universidad de Concepción. Para la cuantificación es importante obtener células únicas en solución. Para ello, las células fueron cuidadosamente lavadas dos veces con PBS 0.6X, tripsinizadas y recuperadas con 200 μ L PBS 0.6X/10% de suero. Las células fueron lentamente resuspendidas en tubos de citómetro y preparadas para la cuantificación de fluorescencia *in vivo*.

Antes de iniciar la cuantificación es recomendable llevar a cabo la compensación del equipo, la cual sirve para corregir la existencia de solapamiento entre los espectros de emisión de los fluorocromos empleados. Para dicho proceso se requieren células negativas (transfectadas sin plásmido), células electroporadas solo con el plásmido CMV-GFPn (verde) y células electroporadas sólo con el plásmido CMV-mCherry (rojo). Esto permitió seleccionar los límites de emisión que correspondían al rojo y al verde. El proceso de compensación resumido implicó pasar por el citómetro una condición de células no marcadas (osteoblastos negativos). Primero se ajustó el voltaje de FSC (Tamaño) y SSC (Complejidad) y el umbral (threshold) a

10.000 para eliminar el *debris*. Posteriormente se pasaron las células que poseían un solo color y se regularon los voltajes para los filtros de GFP y mCherry con los siguientes parámetros de voltaje: FSC (220), SSC (147), GFP (295), mCherry (370). Se midieron células marcadas solo con verde (CMV-GFPn) y células solo con rojo (CMV-mCherry). Así, se escogieron los límites de lo que correspondía a células negativas, rojas y verdes.

Posteriormente se pasaron las células por el citómetro y se dibujaron los cuadrantes que ofrecen una representación biparamétrica a través de un diagrama de puntos. Estos gráficos permiten una representación simultánea de dos parámetros X e Y para un conjunto de células. X representa la señal de un parámetro, en todos los gráficos el eje de las X fue utilizado para definir las células positivas a señal de mCherry (rojo) (Figura 8A) y el eje Y la señal del parámetro GFP (verde) (Figura 8B). De tal forma que aquellas células positivas para ambos parámetros fueron sobrepresentadas en una diagonal, por ejemplo, cuando se transfectó un plásmido dual CMV:GFP y CMV:Cherry (Figura 8C).

Finalmente, todas las muestras fueron grabadas utilizando una velocidad lenta y un pocillo p24 permitió recuperar hasta 5000 eventos. Todos los experimentos cuantificados por citómetro se realizaron por duplicado y al menos en dos réplicas biológicas.

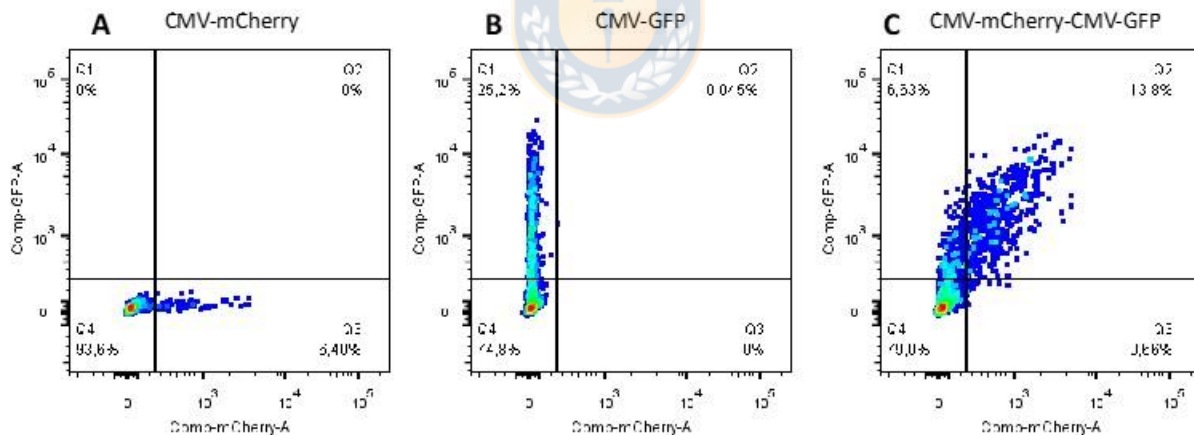


Figura 8: **Compensación de la electroporación**

El proceso de compensación permite corregir el solapamiento entre los espectros de emisión de los fluorocromos empleados. Es muy importante proveer al software de una señal negativa y las señales positivas respectivas. En A, la señal positiva derivada de transfectar un plásmido CMV:mCherry puede detectarse (después de la compensación) en el eje de las Y mientras que en B, células transfectadas con un plásmido CMV:GFP son detectadas únicamente en el eje de las X. En C, se muestra el resultado de transfectar un plásmido positivo tanto para mCherry como GFP, lo que genera una diagonal con los diferentes eventos positivos para ambas marcas. Los cuadrantes Q1-Q4 muestran el porcentaje de células correspondiente a cada población celular. Por ejemplo, la eficiencia de transfección por electroporación de osteoblastos para mCherry fue de 6.4%, para GFP de 25.2% y para el plásmido dual de 13.8%. Las

diferencias entre los valores pueden deberse a varios factores al momento del proceso de electroporación que van desde la calidad de purificación y tamaño del plásmido, tiempo invertido en el proceso de electroporación, calidad y viabilidad celular, etc.

5.5.5 Análisis de datos de citometría

Del citómetro de flujo es posible exportar los datos de fluorescencia de cada una de las células para su análisis estadístico. Los archivos se generan con la extensión .fcs para cada una de las muestras procesadas. Dichos archivos poseen toda la información de la lectura efectuada por el equipo. Los archivos .fcs pueden ser abiertos con el software FlowJo v.10. Primero fue necesario cargar los archivos en el programa y se ajustaron los parámetros para identificar las células que expresan las proteínas fluorescentes. Para esto, se abrió el archivo denominado “negativo” (células electroporadas sin plásmido) que desplegó un gráfico de dispersión, donde cada punto corresponde a un evento cuantificado por el equipo. Se ajustaron los ejes de la siguiente forma: $X = \text{FSC-A}$; $Y = \text{SSC-A}$. De esta manera fue posible analizar los eventos por Tamaño vs Complejidad. El siguiente paso fue elegir la población que corresponde a las células experimentales, eliminando cualquier ruido del equipo o debris. Para esto se crearon diferentes *gates*. Para cada *gate* se buscó la población celular más abundante o la población que mostraba el mayor número de eventos. Después de hacer el *gate* fue posible observar únicamente la nueva población de osteoblastos. Al seleccionar el *gate* creado se despliega un nuevo gráfico con la población elegida. Se ajustaron los ejes $X = \text{FSC-A}$; $Y = \text{FSC-H}$. De esta manera fue posible eliminar células no únicas o “dobletes”. Aquí se realizó un *gate* que dejó fuera la población que se escapó del grupo de eventos de interés. En un nuevo *gate* se trabajó la población elegida, se ajustaron los ejes $X = \text{Comp-mCherry-A}$; $Y = \text{Comp-GFP-A}$ y se colocaron los cuadrantes.

La primera muestra que se analizó fue la negativa (células electroporadas sin plásmido), en la que se obtuvo un cuadrante que permitió discriminar el rojo y el verde (mCherry y GFP). Una vez determinado los límites de los cuadrantes se volvió a la pestaña inicial, se seleccionaron todas las poblaciones que se habían creado y se arrastraron sobre la pestaña *All samples*. Esto permitió que los *gates* y cuadrantes jerarquizados se copiaran en todas las muestras analizadas. Después de aplicar los cuadrantes a todas las muestras fue posible extraer los datos para cada una de las células. La extracción de datos fue posible utilizando el software R.

5.5.6 Análisis estadístico

Para la extracción de datos en R fue necesario descargar el paquete de datos flowCore, cambiar el directorio a la carpeta de archivos.fcs y entrar al la librería del paquete. Posteriormente en R se escribió el script para la extracción de datos.

Debido a que los datos de fluorescencia no se ajustan a una curva de distribución normal fue necesario aplicar métodos estadísticos no paramétricos. Para todos los ensayos de reporteros se utilizó la prueba estadística de Kruskal-Wallis. Primero se convirtieron los archivos .txt a .csv. Se abrieron los archivos y se colocó punto y coma (;) al comienzo y se reemplazaron todos los puntos (.) por comas (,). Posteriormente se cambió la extensión del archivo .txt a .csv. Se abrieron los archivos .csv y se dividieron los datos de fluorescencia de cada célula para llevar a cabo un proceso de normalización. La razón GFP/mCherry permitió tomar en cuenta únicamente valores de fluorescencia de aquellas células transfectadas positivas para el control interno mCherry.

Cada condición fue extraída a diferentes blocs de notas. Se eligió como directorio en R la carpeta donde se guardaron los blocs de notas con los valores GFP/mCherry y se procedió a realizar los diferentes gráficos y análisis estadísticos. La función boxplot permitió graficar en cajas. Para el análisis estadístico fue necesario descargar el paquete **PMCMR**. Posteriormente para el análisis estadístico se escribió en R el siguiente comando:

```
library (PMCMR)  
posthoc.kruskal.dunn.test (c(C1,C2,C3...n), p.adjust="bonf")
```

Con estas líneas se realizó el análisis estadístico de kruskal-wallis. Se cambió el test para ajustar el valor de P. Para todos los casos se ajustó con el test de comparaciones múltiples denominado Bonferroni. En este procedimiento se fijó un nivel de significancia α que se reparte entre cada una de las comparaciones consideradas y se utiliza la desigualdad de Bonferroni. Este test se basa en la creación de un umbral por encima del cual la diferencia entre las dos medias será significativa y por debajo del cual esa diferencia no será estadísticamente significativa. Este método suele ser bastante conservador y se utiliza cuando son muchas las comparaciones a realizar y además, los grupos son homogéneos en varianzas. Tal es el caso de la comparación de medias en ensayos reporteros, lo que ofrece una declaración de significancia mucho más restringida. Finalmente, todos los resultados fueron graficados con *boxplots* y *beanplots* utilizando el paquete de R **ggplot**.

5.6 Ensayo de Captura de Conformación de Cromosoma Circular (4C-seq)

Comúnmente denominada “uno vs todos”, la metodología de 4C-seq permite identificar la frecuencia con la que un fragmento de cromatina interactúa con el resto del genoma en una muestra determinada. Por lo tanto, si se define un fragmento de estudio o “*viewpoint*” en el promotor de un gen, los resultados generarán la frecuencia con la que este promotor interactúa con enhancers u otros módulos *cis*-reguladores del genoma. Esta información es crucial para estudiar el posible paisaje regulatorio de un gen.

Esta técnica requirió varios procesos de estandarización. En esta sección se detallarán el diseño de partidores y la preparación de la muestra, así como las posteriores digestiones y ligaciones necesarias para obtener las librerías de secuenciación. La estandarización fue realizada con base en los protocolos desarrollados por varios laboratorios, particularmente revisados los protocolos de (Brettmann et al., 2018; Göndör et al., 2008; Krijger et al., 2020; Matelot & Noordermeer, 2016; Stadhouders et al., 2013). Todos estos protocolos habían sido aplicados en células de mamífero, principalmente ratón y humano. Por lo tanto, se hicieron modificaciones puntuales para aplicar la técnica correctamente en *Xt*.

Todos los pasos se estandarizaron para 10 millones de células. En 4C se remienda entre 5 y 10 millones, la razón del gran número es que las topologías del genoma son variables entre células y entre alelos individuales de células únicas. Cada célula diploide sólo puede contribuir con un máximo de dos fragmentos de ADN ligados proximalmente (1 por cada alelo), por lo tanto, para obtener el perfil de contactos de un *viewpoint*, la PCR inversa necesita ser realizada en el equivalente de miles de genomas, lo que genera perfiles de contactos reproducibles que exponen la mayor parte de conformaciones predominantes en una población celular (Krijger et al., 2020).

5.6.1 Selección de enzimas

Se realizó una búsqueda de cada promotor de los genes de interés en el ensamble del genomeoma de *X. tropicalis* Jul. 2016 (*Xenopus_tropicalis_v9.1/xenTro9*) del sitio UCSC Genome Browser <http://genome.ucsc.edu/>. Se estudiaron las posibles combinaciones de enzimas comúnmente usadas en ensayos de captura de conformación de cromosoma que al digerir generaran fragmentos “*viewpoints*” sobre las regiones promotoras y que a su vez fueran compatibles con una segunda enzima que acorta cada fragmento para generar los “*baits*” sobre los cuáles se diseñarían los partidores para la PCR inversa. Un *viewpoint* puede abarcar exactamente la zona de interés (por ejemplo, un promotor) o estar en la vecindad de ella (~1 Kb).

Las enzimas seleccionadas en un primer diseño fueron *DpnII* y *HaeIII* como primera y segunda enzima respectivamente. Al respecto, es muy importante cerciorarse que las enzimas seleccionadas son compatibles no sólo virtualmente, sino en la aproximación experimental ya que no todas son óptimas para digerir cromatina, en nuestras manos comprobamos que la eficiencia de digestión depende de varios factores tales como la preparación de la muestra, el tipo celular y la especie.

En las primeras pruebas de estandarización la enzima *DpnII* degradó la cromatina de osteoblastos de *Xt* (Figura 8^a), lo que fue perceptible en el proceso de ligación ya que en una cromatina degradada la eficiencia de ligación disminuye drásticamente. Como control de estandarización se intentó digerir con la misma enzima células *Hek293*, lo que generó una digestión óptima.

Al observar esta dificultad, una primera opción fue cambiar la enzima *DpnII* por su esquisoizómero *MboI*, una enzima comúnmente usada en ensayos de HiC. El buffer CutSmart mantuvo la cromatina en condiciones óptimas (Figura 8B), sin embargo, la digestión fue sólo parcial (Figura 8C). Las posibilidades con otras enzimas que cortaran en el promotor eran mínimas y llevaban a utilizar varias preparaciones de osteoblastos, lo que no era adecuado por la dificultad de cultivar millones de células primarias. Para estar seguros que la digestión parcial era sólo para el caso de *MboI* se intentó digerir cromatina de osteoblastos con las enzimas *HindIII* y *HaeIII*, las cuáles también resultaron inevitablemente en digestiones parciales.

Lo anterior llevó a revisar el buffer de lisis de osteoblastos ya que las características del tipo celular podrían favorecer negativamente la eficiencia afectando la subsecuente digestión de la cromatina. Al no existir hasta el momento protocolos aplicados al tipo celular de nuestro interés buscamos combinaciones de buffers cambiando la concentración de SDS en cada uno de ellos. Finalmente, el uso de un buffer de lisis hipotónico con 0.5% de SDS fue óptimo para la digestión con *MboI*. Dicho buffer también fue comprobado con varias enzimas, por ejemplo, *HindIII* mostró un *smear* aproximadamente a 5 kb, lo que corresponde a una eficiente digestión de enzimas que cortan cada 6 pb mientras que *MboI* y *Csp6I* cortan en 4 pares de bases con un *smear* entre 0.5 y 1 kb, lo que indicó una óptima digestión (Figura 8D).

En conclusión, se utilizó la enzima *MboI* para la primera digestión debido a que era la que liberaba los fragmentos de todos los *viewpoints* en una sola preparación de 10 millones de osteoblastos. Además, dicha enzima generaba fragmentos que eran factibles de digerir con la segunda enzima *Csp6I* que liberaba todos los *baits* de interés. Ambas enzimas son de alta frecuencia de corte (4 pares de bases) y han sido utilizadas en ensayos de conformación de cromosoma.

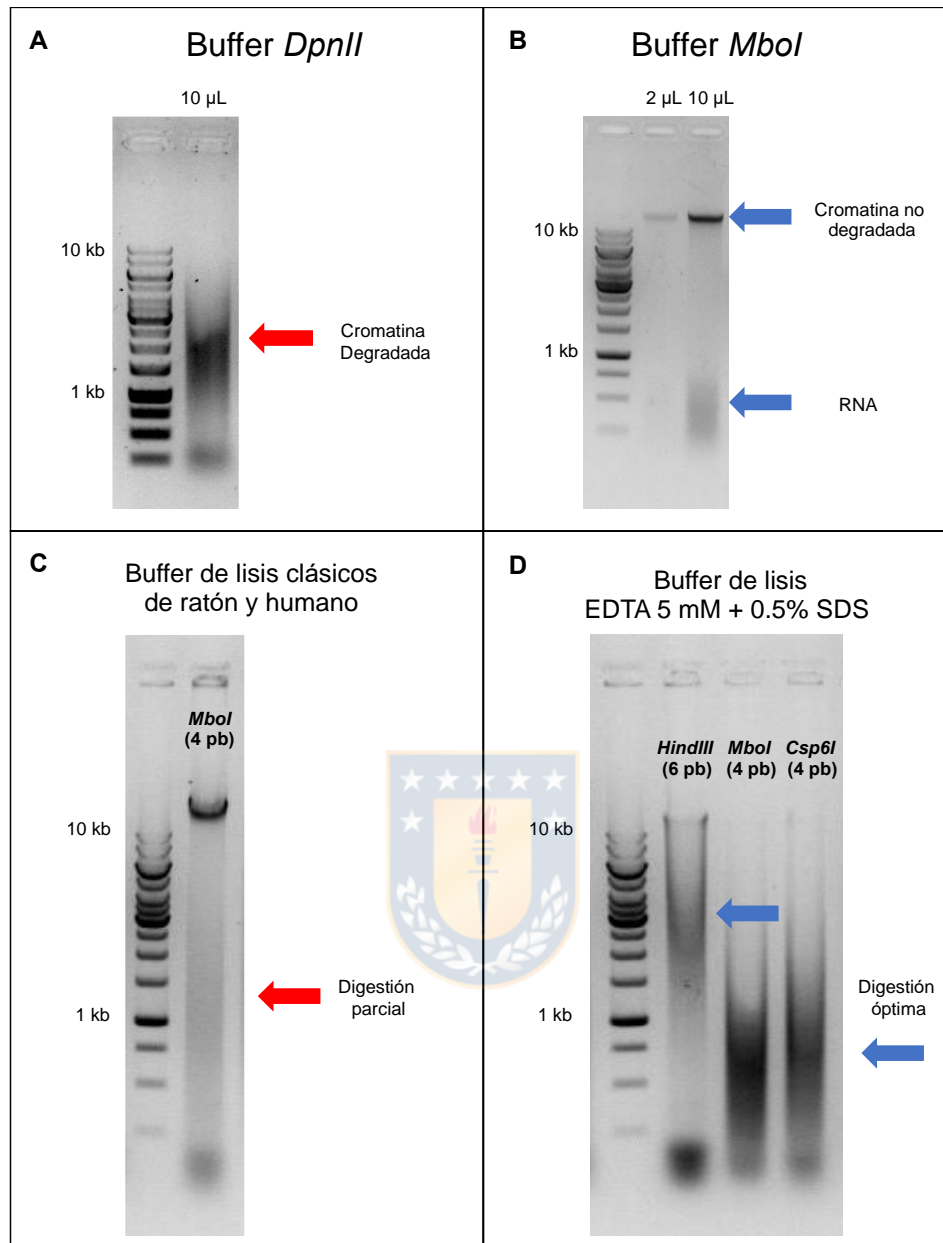


Figura 9: **Estandarización de la primera digestión**

En A se observa el resultado de una cromatina degradada por el buffer de la enzima *DpnII* en osteoblastos lisados con buffer clásicos aplicados en ensayos 4C de mamífero. En B, la enzima *Mbol* (isoesquizómero de *DpnII*) utiliza un buffer *CutSmart* que mantiene la cromatina en condiciones óptimas. En C, aunque el buffer de *Mbol* mantiene la cromatina estable, la digestión sólo fue parcial (con otras enzimas el efecto fue similar). En D, el uso de un buffer de lisis hipotónico con 0.5% de SDS fue óptimo para la posterior digestión, lo cuál fue comprobado con varias enzimas.

5.6.2 Diseño de partidores

Se buscó amplificar regiones o *viewpoints* (*Vp*) que abarcaran el TSS de cada gen para evaluar su frecuencia de contacto con los enhancers putativos. En la figura 8 se muestra el diseño de partidores para 4C usando como ejemplo el promotor del gen *SCPPA2*. El tamaño de un *Vp* debe ser mayor a 400 pb y menor a 5 kb, lo cuál depende de los sitios de corte de la primera enzima de restricción seleccionada (R1). Los fragmentos son digeridos con una segunda enzima (R2) que reduce el tamaño del *Vp*. Los partidores se diseñaron al final del 5' y 3' del fragmento *Vp* delimitado por la primera enzima (Figura 8A). El fragmento final (después de ser procesado con la segunda enzima) se denomina *bait* y preferentemente debe ser mayor a 200 pb para asegurar un espacio suficiente para el diseño de partidores, así como una eficiente circularización después de la segunda ligación. La secuenciación de las librerías 4C inician en el sitio R1, por lo que el partidador que flanquea este sitio contendrá el adaptador illumina P5. Este “reading primer” se diseña preferentemente sobreponiéndose con el sitio de la enzima R1 o lo más próximo a él para asegurar la lectura de secuencias de mayor tamaño que interactúan con el *Vp*. El partidador P7 “no reading” puede ser diseñado en la proximidad del sitio R2 y permite una mayor flexibilidad. Inherente a la estrategia experimental del 4C, dos fragmentos serán amplificados con mayor frecuencia, el fragmento “no digerido” (número 1), debido a la falla de corte del sitio R1 cerca del reading primer y cuyo producto puede trazarse (para este ejemplo, aproximadamente 150 pb) y el auto-círculo (número 2) del *vp* en la primera ronda de ligación (Fragmento R1-R2 en el lado opuesto del fragmento *vp* no usado para el diseño de primers). Si el tamaño de estos fragmentos es tan corto para circularizar o tan largo para amplificar, los productos no siempre estarán presentes (Krijger et al., 2020; Stadhouders et al., 2013). La imagen de la figura 8B muestra los sitios de restricción de las enzimas *MboI* y *Csp6I*, el primero y segundo exón del gen *SCPPA2* así como la matriz de conservación del promotor (verde) visualizado en el sitio web UCSC Genome Browser.

Para el diseño de todos los partidores se utilizó el software libre *Primer3Plus* y la herramienta Primer designing del NCBI–NIH. Todas las secuencias de partidores para los diferentes *bait* se encuentran en el anexo 1.

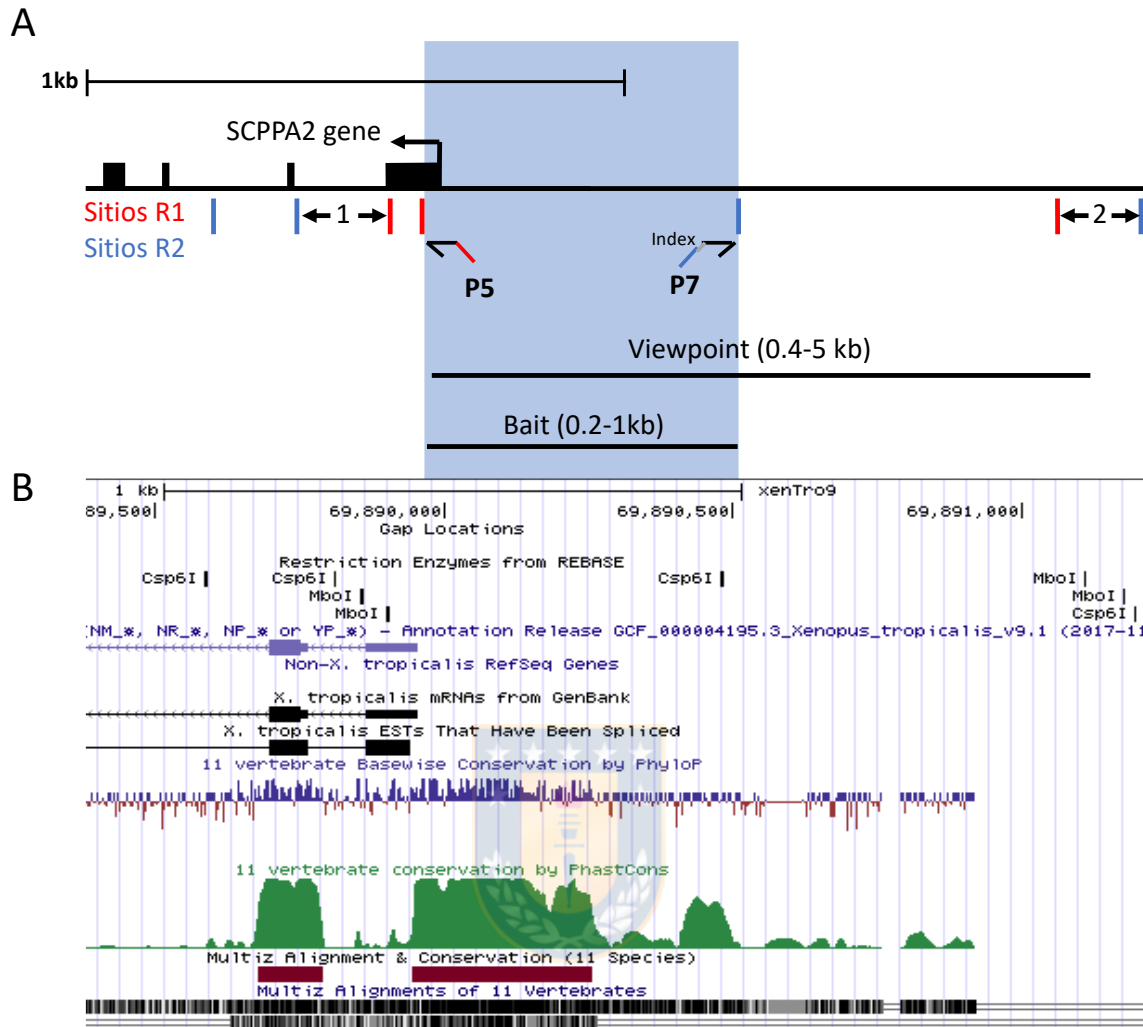


Figura 10: Ejemplo del diseño de partidores para 4C

En A se esquematiza una ventana de ~1.6 Kb dibujada a partir de la imagen en B (UCSC Genome Browser). Los sitios donde corta la enzima R1 están marcados en rojo y en azul la enzima R2. El fragmento *viewpoint* después de la primera digestión incluye el TSS del gen *SCPPA2* y libera un fragmento aproximado de 1 Kb (zonas gris + amarillo). Para seleccionar la primera enzima es importante que abarque el MCR de interés, que sea mayor a 0.4 Kb y menor a 5 Kb. La segunda enzima corta el *viewpoint* en dos y libera el *bait*, el cuál incluye la región o MCR de interés. En los sitios de corte de R1 y R2 se diseñan partidores que en una PCR inversa amplificarían los fragmentos capturados por el *bait*. A dichos partidores se les agrega un adaptador de la plataforma de secuenciación *illumina* (P5 y P7). Entre el adaptador y el primer P7 se coloca una etiqueta o *index* que permite separar el resultado de secuenciación de diferentes *viewpoints*. En B, se muestra la alta conservación (en verde) del exón 2 y de la región promotora del gen *SCPPA2*, esta última incluida en el *bait* así como los diferentes sitios de corte de las enzimas R1 y R2.

5.6.3 Amplificación y ligación de *Viewpoints*

Para probar la eficiencia de los partidores diseñados en el punto anterior se realizó la amplificación y clonamiento de todos los *bait*s de los diferentes *viewpoints* en el vector *pBlueScript*. Cada fragmento fue amplificado utilizando la enzima polimerasa KOD (Calbiochem, San Diego, CA, USA), siguiendo las instrucciones del fabricante. Los productos PCR fueron digeridos para tener la certeza de que los sitios de corte de la enzima R1 (*Mbol*) efectivamente se encontraban en el genoma de *Xt* y no han sido modificados por alguna mutación puntual. Un ensayo de restricción permitió corroborar que todos los sitios de la primera enzima estaban presentes en los *viewpoint* seleccionados (Figura 9A). Posteriormente, los productos de PCR fueron ligados a vectores *pBluescript* utilizando 30 ng del vector y 90 ng de inserto y 1 μ L de enzima ligasa (Invitrogen) durante 2 h a 37°C. Los plásmidos resultantes fueron utilizados como templado para realizar una PCR inversa. Los partidores fueron diseñados en cada extremo del *bait*, por lo que la PCR amplificó el cuerpo del plásmido *pBlueScript* de aproximadamente 3 Kb (Figura 9B)

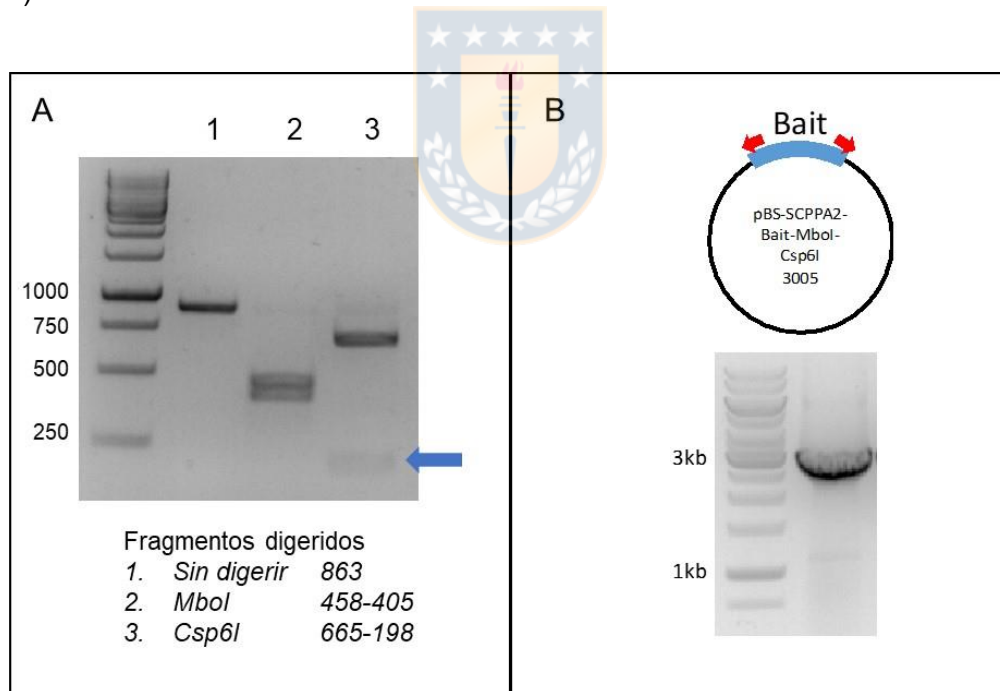


Figura 11: **Amplificación de *viewpoints***

En A, un fragmento que contiene el *viewpoint* del gen *SCPPA2* fue digerido con las enzimas diseñadas para el 4C. La enzima *Mbol* liberó dos fragmentos de tamaños cercanos, mientras que *Csp6I* liberó un fragmento de 665 pb y uno de 198 pb (flecha azul). Lo que sugiere que ambos sitios se encuentran en el fragmento diseñado como *viewpoint*. En B, el *viewpoint* es ingresado a un vector que es usado como templado en una PCR inversa que correctamente amplifica el *backbone* del plásmido.

5.6.4 Preparación de la muestra

El protocolo fue aplicado en células de cultivo primario de osteoblastos al 90% de confluencia (Figura 10A). El método para obtener las células de encuentra detallado en la sección 5.2. En el pasaje 2 se colectaron 12 millones de células que fueron tripsinizadas, lavadas y colectadas en un tubo *falcon* con 12 ml de medio de crecimiento de osteoblastos y se evaluó su óptima viabilidad (Figura 10B). La razón de partir con 12 millones es que este protocolo requiere de varias centrifugaciones, por lo que durante el proceso se pierden células. Idealmente, el proceso de entrecruzamiento o *crosslinking* está estandarizado para 10 millones. Las células fueron centrifugadas (160 rcf) durante 5 min y se descartó el sobrenadante. El pellet fue resuspendido cuidadosamente en 6 ml de MCO.

5.6.5 Crosslinking

El crosslinking se llevó a cabo a una concentración final del 2%, para lo cuál se agregaron 6 ml de formaldehído 4% (16% formaldehyde solution, methanol-free Thermofisher 28906) a la muestra y se incubó durante 10 min a TA agitando a 90 rpm en un agitador vertical. Para detener el entrecruzamiento se agregó 1.7 ml de glicina fría 1M (concentración final 125 mM o 1/8 del volumen final). Se centrifugó 10 min / 650 rcf / 4°C. Posteriormente se lavó el *pellet* mediante una cuidadosa resuspensión en 12 ml de PBS frío. Se centrifugó la muestra 10 min / 650 rcf / 4°C y se removió el sobrenadante sin tocar el pellet. Se agregó 1 ml de PBS frío y se dividieron las células en dos tubos *ependorf* de 1.5 ml previamente etiquetados y enfriados en hielo (~5 millones de células cada uno). Finalmente, se centrifugaron 10 min / 650 rcf / 4°C, se removió el sobrenadante y se congelaron las muestras inmediatamente en nitrógeno líquido.

5.6.6 Lisis celular

Se descongelaron ambos tubos en hielo (30 min-1h) y se resuspendieron cada uno en 500 µL de EDTA 5 mM. 5 µL de cada tubo fueron colocados en un tubo *ependorf* de 1.5 ml (~100,000 células) y almacenados a 4 °C. El resto se centrifugó 5 min / 650 rcf / TA y se descartó el sobrenadante. Cada *pellet* fue resuspendido en 500 µL de EDTA 5 mM. Se centrifugó 5 min / 650 rcf / TA y se descartó el sobrenadante (mientras tanto se preparó un thermomixer a 65 °C). Las células fueron resuspendidas cuidadosamente en 100 µL de buffer de lisis (EDTA 5 mM con 1X

de inhibidor de proteasas) preparado previamente, se llevaron a un volumen a 122 μL y se agregaron 3.1 μL de SDS al 20% (volumen final 125 μL). Cada tubo se mezcló por inversión 5 veces y se centrifugó 30 s. Las muestras fueron incubadas en el thermomixer por exactamente 10 min a 65 °C / 900 rpm seguido por incubación toda la noche a 37 °C en agitación a 900 rpm. Para asegurarse que la lisis estaba completa se mezclaron 5 μL de cada muestra más 10 μL de azul de tripano (total 20 μL) en un tubo de 0.6 ml. Se utilizaron 6 μL de la mezcla en cámara de Neubauer y el resto se tiñó con DAPI (1:100) más Faloidina (1:100). Se realizó el mismo procedimiento con la alícuota previamente almacenada a 4 °C para comparar “no lisadas” contra “lisadas”. Mientras se observaba la eficiencia de la lisis en microscopio de campo claro incubó 20 min el resto de la suspensión para documentar posteriormente la eficiencia de la lisis por microscopia de fluorescencia. En una lisis ineficiente las células pueden aparecer blancas y con contornos brillantes (Figura 10C) mientras que en una lisis eficiente se muestra el interior celular de color azul oscuro (Figura 10D). Una mejor evaluación de la lisis puede realizarse con tinción con DAPI y faloidina. En una lisis ineficiente pueden apreciarse células íntegras con abundante citoplasma marcado con faloidina (Figura 10E y F) mientras que en la lisis eficiente sólo se aprecian los núcleos íntegros con escasos restos citoplasmáticos marcados con faloidina (Figura 10G y H).



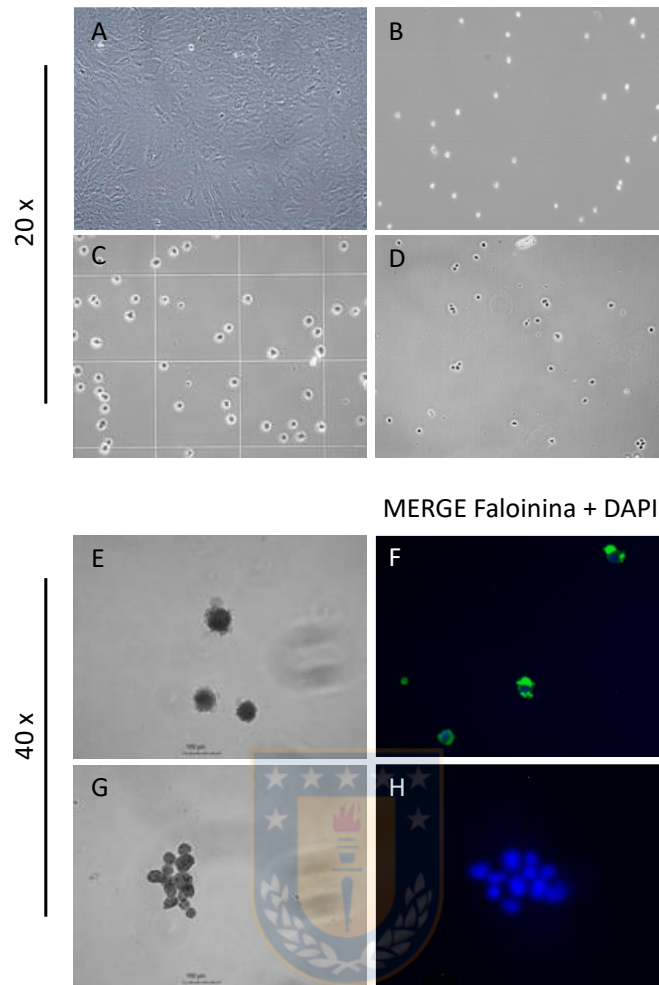


Figura 12: **Optimización de la lisis celular**

En A se muestran células primarias de osteoblastos en pasaje 2 adheridas con una confluencia mayor al 90%. Es muy importante que las células estén limpias, sin grumos ni gránulos de estrés para una correcta fijación y subsecuente lisis. En B las células tripsinizadas y teñidas con azul de tripano con una viabilidad del 100% (células que presentan abundantes gránulos o baja viabilidad en esta fase deberían ser descartadas para experimentos de 4C). En C células fijadas con formaldehído, lisadas y teñidas con azul de tripano muestran una ineficiente lisis debido al uso de buffers no compatibles con osteoblastos. En D, células lisadas al 100% con un buffer hipotónico (EDTA + SDS) muestran un color azul oscuro intenso y son aptas para digerir con la enzima *Mbol* (células lisadas parcialmente no son adecuadas para digerir y deberían tratarse más tiempo en buffer de lisis, no es recomendable el uso de homogeneizador, pero se puede realizar un pipeteo suave para mejorar la eficiencia al disgregar los conglomerados celulares). En E, células fijadas no lisadas muestran una estructura con abundante citoplasma, el cuál se puede visualizar en F debido al marcaje con faloidina (1:100) y DAPI (1:1000). En G (campo claro) y H (doble marcaje, faloidina y DAPI) las células lisadas correctamente solo contienen núcleos con escasos o nulos restos citoplasmáticos (ausencia de marcaje de faloidina).

5.6.7 Primera digestión

Para la primera digestión se preparan previamente buffer de lisis fresco con SDS al 0.5% final y buffer de digestión (210 μ L de buffer (10x), 189 μ L de tritón 1.8% (20%) y 826 μ L H₂O). El buffer se utiliza de acuerdo a la enzima seleccionada. Para *Mbol* se utilizó el buffer *CutSmart* (NEB). Se juntaron ambos tubos de células lisadas (250 μ L), se agregaron 625 μ L de buffer de lisis y se mezcló por inversión. Se dividió la mezcla en 7 tubos (125 μ L cada uno) y se agregó a cada muestra 175 μ L de buffer de digestión fresco (300 μ L de reacción total). Se incubó 1h / 37°C / 900 rpm. Se apartaron 2 μ L de cada reacción (14 μ L totales) como “control no digerido” y se almacenaron a 4°C. Para la digestión se agregaron 4 μ L (100 U) de enzima (*Mbol* enzyme (25U/ μ L) NEB R0147M) a cada muestra y se incubó 4 h / 37°C / 900 rpm. Nuevamente se agregó 4 μ L (100 U) de enzima *Mbol* a cada muestra y se incubó toda la noche / 37°C / 900 rpm. Al día siguiente se evaluó la eficiencia de la digestión y se llevó a cabo un *checkpoint* para evaluar la eficiencia de la digestión.

Para el *ckeckpoint* de la etapa de restricción se colectaron 3 μ L de cada reacción y se juntaron en un tubo de 1.5 ml (21 μ L total), se agregaron 76.5 μ L de 10 mM Tris-HCl pH 7.5 y 2.5 μ L de proteinasa K (20 mg/mL) y se incubaron 2 h at 65 °C. Se agregaron 100 μ L de fenol-cloroformo-alcohol isoamílico 25:24:1 Sigma 77617 y se mezcló varias veces por inversión (no vortex). Se centrifugó 5 min 16100 rcf a TA. Se transfirió la fase acuosa a un nuevo tubo y se agregaron 6.66 μ L de acetato de sodio 3M pH 5.2, 1 μ L de glicógeno 20 mg/mL, y 300 μ L de 100% etanol (EtOH). Se mezcló por inversión y se enfrió 1 h a -80 °C. Se centrifugar 20 min / 16100 rcf / 4 °C se removió el sobrenadante y se agregó 70% EtOH para lavar el *pellet*. Se centrifugó 5 min / 16100 rcf / 4 °C, se removió el sobrenadante y se hizo un breve *spin*. Se retiraron los residuos de EtOH con pipeta p10 y se secó el pellet 10 min a TA. Se resuspendió el *pellet* en 20 μ L of H₂O libre de nucleasas, se cargaron 10 μ L en un gel de agarosa 0.8 % y se realizó una electroforesis a 100 V / 40 min. Una banda mayor a 10 kb fue visible (Figura 11A) en el control “no digerido” y un claro *smear* alrededor de 1 kb en el control “digerido”. Cuando la digestión fue óptima se llevó a cabo el proceso de ligación.

5.6.8 Primera ligación y purificación del templado (3C)

Se colocaron las siete reacciones a 65°C / 20 min para inactivar la enzima *Mbol*. Mientras tanto, se preparó una mezcla de ligación en hielo utilizando la enzima T4 DNA HC Ligase (Thermofisher

EL0013). La mezcla se realizó para 7.2 volúmenes utilizando 878.4 de H₂O_mQ, 360 µL (1X) de buffer de ligación (10x), 180 µL (0.5 mM) de ATP (10 mM) y 23.8 µL (100 U) de ligasa (30U/µL).

Las reacciones fueron colocadas inmediatamente en hielo, se agregaron 200 µL de la mezcla de ligación a cada reacción (Volúmen final 500 µL) y se incubaron 22 h / 16°C / 300 rpm. Se realizó un *ckeckpoint* (eficiencia de ligación) siguiendo los pasos del apartado anterior. Cuando la ligación fue óptima se centrifugaron todas las reacciones a 650 g / 15 min / TA. Cuidadosamente se removió el sobrenadante y se resuspendieron los núcleos de cada tubo en 43 µL de buffer TE. Se juntaron todas las muestras en un solo tubo de 1.5 ml (301 µL totales).

Se llevó a cabo un proceso de *de-crosslink* utilizando 10 µL de proteinasa K (6U) (Thermofisher EO0491) a la mezcla total y se incubó toda la noche a 65°C. Para la purificación se enfrió la reacción a 37°C, se agregaron 5 µL de RNAse (7.5 U/ µL) Roche 1119915 y se incubó 30 min / 37°C. Se transfirió la muestra a un tubo *Phase Lock Gel* (PLG) de 2 ml y se agregaron 316 µL de fenol: cloroformo : alcohol isoamílico. Se mezcló por inversión exhaustivamente y se centrifugó 10 min/12,600 rcf / TA. La muestra fue transferida a un nuevo tubo PLG de 2 ml y se agregaron 310 µL de cloroformo. Se mezcló por inversión exhaustivamente y se centrifugó 10 min / 12,600 rcf / RT. Se transfirió la muestra a un tubo *ependorf* de 1.5 ml, se agregaron 30 µL de 3M de acetato de sodio (NaOAc) Invitrogen AM9740 y 1 µL de glicógeno Roche 10901393001. Se mezcló por inversión y se agregaron 900 µL etanol al 100% (concentración final de EtOH 75%). Se mezcló exhaustivamente y se incubó a -20°C toda la noche. Posteriormente se centrifugó la muestra 30 min / 21,000 rcf / 4°C y cuidadosamente se descartó el sobrenadante. Se lavó el *pellet* con 1 ml de etanol frío al 70% y se centrifugó 5 min / 21,000 rcf / 4°C. Se lavó el *pellet* con 1 ml de etanol frío al 70%, se centrifugó 5 min / 21,000 rcf / 4°C, se removió el etanol completamente. Se secó el *pellet* a TA (10 min) y se resuspendió en 50 µL de H₂O libre de nucleasas. Finalmente, 1 µL de ADN fue resuspendido en 9 µL de H₂O libre de nucleasas para su cuantificación por *Qubit dsDNA BR Assay kit Invitrogen Q32850* (199 µL de BR buffer + 1 µL sonda + 1 µL de ADN diluido) y se almacenó el templado ADN (3C) a -20°C hasta continuar con la segunda digestión.

5.6.9 Segunda digestión

La segunda digestión se llevó a cabo utilizando 11.5 µg de ADN (3C) con 50 U de enzima *Csp6I* (25U/µL) Thermofisher ER0211. Para la digestión se sugiere una concentración máxima de 100 ng/µL. La concentración final se obtuvo con el cociente de 11500 ng / 200 µL = 57.5 ng/µL. La

mezcla se compuso de 125 μ L de agua mQ, 20 μ L (1X) de buffer (10x), 50 μ L (11.5 μ g) de ADN, 5 μ L (50 U) de enzima (10 U/ μ L) para una reacción total de 200 μ L.

La reacción fue incubada toda la noche a 37°C / 500 rpm. Se determinó la eficiencia de digestión utilizando 100 ng de reacción digerida en un volumen de 15 μ l con 10mM Tris-HCl pH 7.5. Se agregaron 1.5 μ l de buffer de carga y se realizó una electroforesis en gel de agarosa 0.8 % a 100 V / 45 min. El ADN digerido mostró un claro *smear* por debajo de 2.5 kb lo que consiste en una digestión óptima.

5.6.10 Segunda ligación y purificación del templado (4C)

La reacción de segunda ligación se realizó a una concentración final diluida de 1 ng/ μ l. La muestra fue transferida a un tubo de 50 ml. Se inactivó la enzima *Csp6I* 20 min / 65°C. Inmediatamente la reacción fue colocada en hielo y se ensambló con 9570 μ L de H₂O mQ, 200 μ L de ADN digerido con *Csp6I*, 1150 μ L (1X) de buffer de ligación (10x), 575 μ L (0.5 mM) de ATP (10 mM), 5 μ L (150 U) de ligasa (30U/ μ L). La reacción total de 11500 μ L fue incubada 22h / 16 °C .

La purificación fue realizada siguiendo el protocolo de (Brettmann et al., 2018) utilizando 35 mL (75%) de EtOH (100%), 7 μ l de glicógeno, 470 μ l (1:100, 3M, pH 5.5) de Acetato de Sodio y 23 μ L de H₂O mQ (reacción total de 47 ml). Se mezcló y se incubó a -20°C toda la noche. Al día siguiente se centrifugó 45 min / 3900 rcf / 4 °C, se removió el sobrenadante y se agregaron 10 mL de EtOH al 70%. La muestra se centrifugó 15 min / 3900 rcf / 4 °C. Se removió el sobrenadante y se centrifugó 1 min para remover todo el etanol. El pellet fue secado a TA (~20 min), posteriormente se agregaron 150 μ L de 10 mM Tris-HCl pH 7.5 seguido de una incubación a 37 °C hasta disolver el pellet. Cuando el pellet está bien disuelto se transfirió la reacción a un tubo de 1.5 ml y se purificó utilizando el kit *QIAquick PCR Purification* (QIAGEN 28106) con algunas variantes.

Se agregar 750 μ l de buffer PB (5 volúmenes) a la reacción (Total 900 μ l) y se prepararon 3 columnas a las que se les agregó 300 μ l de reacción. Se incubó por 2 min a TA. Se centrifugaron las 3 columnas por 1 min y se descartaron los sobrenadantes. Se agregaron 750 μ l de buffer PE y se centrifugó por 1 min. Se colocaron las columnas nuevos tubos de 1.5 ml y se centrifugaron por 1 min. Se juntó el ADN en un solo tubo (~100 μ l) y se transfirió a hielo. 1 μ l de ADN fue diluido en 9 μ l de buffer de elución para su cuantificación por *Qubit*. Finalmente, la muestra fue rotulada (ADN templado 4C) y almacenada a -20°C hasta la realización de la PCR inversa.

5.6.11 Generación de librerías (4C) por PCR inversa

Se realizaron una serie de PCRs con diferentes diluciones del templado para elegir la concentración óptima de amplificación utilizando la *Expand™ Long Template PCR System* (Sigma 11759060001). Se diluyó el ADN templado a 12.5, 25, 50, 100 ng en 21.25 µL totales de H₂O_mQ y se realizaron las siguientes mezclas de reacción para cada *viewpoint* con partidores forward y reverse sin adaptadores.

	1V	4.2V
H ₂ O _m Q	20 µL	84 µL
Buffer-1 (1x)	5 µL (10x)	21 µL
dNTPs 0.2 mM	1 µL (10 mM)	4.2 µL
Primer iF (25 pmol)	1 µL (25 µM)	4.2 µL
Primer iR (25 pmol)	1 µL (25 µM)	4.2 µL
Templado 4C (12.5, 25, 50, 100 ng)	21.25 µL	-----
DNA polimerasa (5U/ µL)	0.75 µL	3.2 µL
Reacción total	50 µL	120.8 µL

Table 2: PCR inversa

La mezcla de 4.2V fue dividida en tubos PCR. 28.75 µL en cada tubo y 21.25 µL de cada concentración de templado para reacciones de 50 µL. Se programó el termociclador (1 ciclo 2 min-94°C; 30 cliclos 15s-94°C; TM 1min-55°C; 3min-68°C; 1 ciclo 7min-68°C. 20 µL del material fueron visualizados en un gel de agarosa largo al 2% / 90V / 80 min / 1 kb *ladder*. La concentración óptima fue cuando la intensidad del *smear* incrementó linealmente, donde el tamaño de rango de productos PCR fue máxima (1-1.5 kb) y donde mínimos dímeros de partidores fueron detectados (Figura 12).

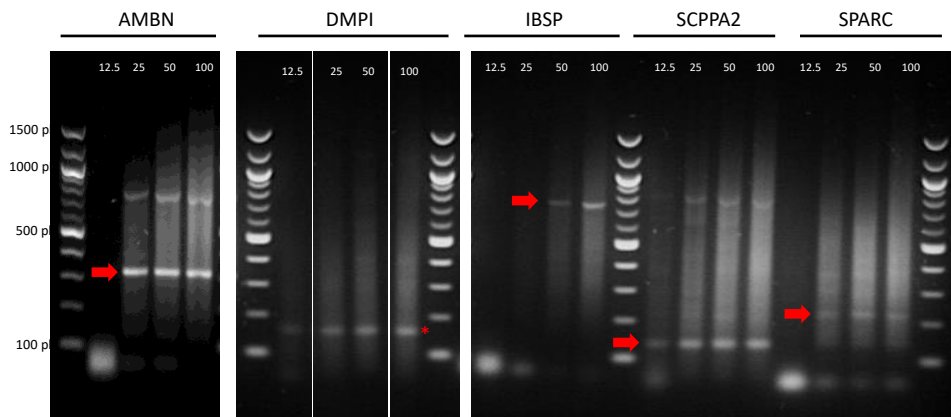


Figura 13: Optimización de PCR inversa

Los 5 baits de estudio fueron analizados mediante un gel de agarosa. A 100 ng la concentración ya no se incrementa al doble, en las PCRs de *Ambn*, *Scppa2* y *Sparc*, 100 ng es similar a 50 ng por lo que a 100 ng podría estar saturado. 50 ng para la mayoría de PCRs incrementa respecto a 25 ng pero, según una apreciación visual, no incrementa al doble. Probablemente en 50 ng se empieza a saturar. Por lo tanto, 40 ng sería una condición óptima para amplificar todos los *Vp*. Es importante notar que el resultado de la PCR frecuentemente muestra un *smear* y un par de bandas de mayor intensidad. Estas dos bandas “no informativas en el mapeo final” son frecuentes en la preparación de librearías y son causadas por 1) El fragmento “no digerido” debido a la falla de cortar en el sitio R1 siguiente del P5 reading primer (el fragmento RE1-RE2 directamente adyacente al fragmento RE1-RE2 del *vp*) y 2) El fragmento “auto-circularizado” (auto-ligación) del *vp* en la primera ronda de ligación (el fragmento RE1-RE2 en el extremo opuesto final del fragmento *vp* no usado para el diseño de partidores). Si el tamaño de estos fragmentos es tan corto para circularizar o tan largo para amplificar, los productos de PCR no siempre son visibles. El tamaño de las bandas de los fragmentos “no digeridos” pueden ser trazados de acuerdo a la distancia entre el sitio *MboI* (Reading primer) y el subsecuente sitio *Csp6I* y sumando ~50 pb (Matelot & Noordermeer, 2016). Para *Ambn* (350 pb); *Dmp1* (1200); *Ibsp* (750); *Scppa2* (150 pb); *Sparc* (250 pb). En todos los casos es posible trazar el fragmento no digerido (flechas rojas), excepto para DMPI, cuya secuencia revela que la banda de 150 pb podría tratarse probablemente del fragmento auto-ligado (asterisco rojo). Nota: Si no hay un *smear* y las bandas (no digerido y/o autoligado) están sobrecargadas, probablemente hubo un problema en la preparación del templado 4C.

Una vez resuelta la concentración óptima, se realizó una PCR con adaptadores P5/P7 para probar los partidores y la concentración óptima del templado. Se cargaron 20 μ L del resultado de amplificación al lado de 20 μ L de la PCR anterior (sin adaptadores). Un *shift* de aproximadamente 100bp fue percibido debido al mayor tamaño de los partidores con adaptadores (Figura 13).

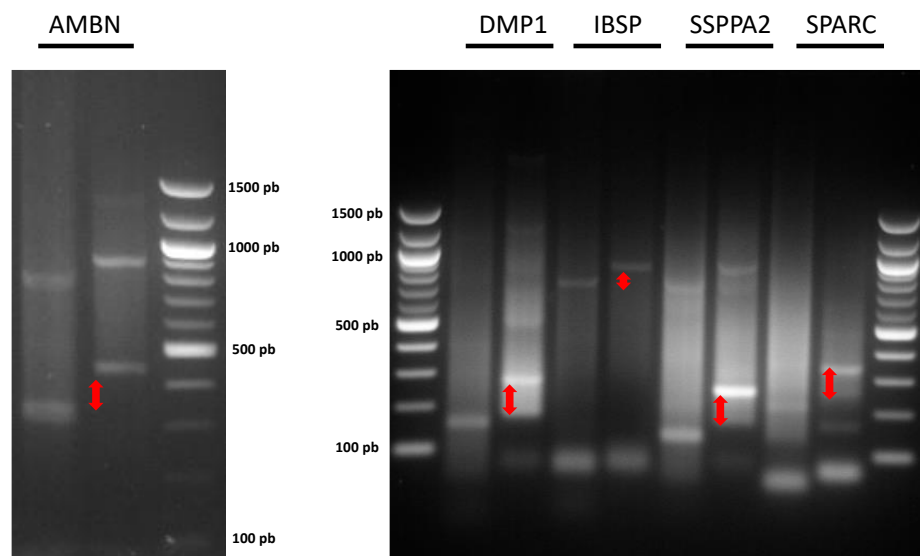


Figura 14: Optimización de PCR inversa con adaptadores

Los 5 baits fueron amplificados con primers P5/P7 con adaptadores utilizando 40 ng como templado. 20 μ L fueron cargados en un gel de agarosa al 2% y el patrón de bandas fue comparado con respecto al producto amplificado de cada *bait* (50 ng) de la figura 3. Observar que las bandas predominantes en el gel migran \sim 100 pb arriba cuando se amplificó con adaptadores P5/P7 debido al mayor tamaño del producto (flechas rojas).

Se realizaron 12 PCRs con 40 ng totales (concentración óptima para este experimento) de templado y partidores con adaptadores para cada *Vp*. Se juntaron todas las reacciones y se purificaron por *QIAquick PCR Purification* utilizando 2 columnas por cada *Vp*. La elución de cada columna fue realizada con 35 μ L de buffer de elución y se juntaron los productos. Posteriormente se utilizaron 50 μ L de cada reacción para purificarlas por *Agencourt AMPureXP Beckman A63881*. El proceso de purificación con este método consistió en temperar al menos 30 min una alícuota de *AMPureXP beads*. Se agitó exhaustivamente antes de usar y se gregan 40 μ L de *AMPureXP beads* a cada muestra. Después de una breve agitación y *spin* las muestras se incuban 5 min a TA. Se colocaron en un rack de separación magnético durante 5 minutos. Una vez que la solución fue clara se agregaron cuidadosamente 500 μ L de EtOH al 80% para lavar los *beads*, 3 minutos después se removió el sobrenadante y se agregaron cuidadosamente 500 μ L de EtOH al 80%. 3 minutos después se removió el sobrenadante y se dio un breve *spin* en microcentrifuga. Se colocaron las muestras en el *rack* de separación magnético y después de 3 minutos se removieron los residuos de sobrenadante con pipeta p10 y se secó el *pellet* 2 min a 37°C. Se resuspendieron los *beads* en 50 μ L de 10mM Tris-HCl pH 7.5 y se incubó cada muestra 10 min / 600 rpm / TA. Después de un breve *spin* en microcentrifuga se colocaron las muestras en el rack de separación magnético. Después de 3 minutos se recuperaron 45 μ L de muestra. Finalmente, cada muestra fue cuantificada por Nanodrop para determinar la pureza. La razón de absorción A260/280 debería estar entre 1.8 y 2 mientras que la razón de absorción 260/230 debería ser mayor a 2.25 De acuerdo a (Krijger et al., 2020). Para conocer la concentración más exacta se cuantificó 1 μ L por Qubit y se corrió un gel con 200 ng de ADN por cada muestra. Cuando los dímeros de *primers* (bandas debajo de 200 pb) no aparecieron en el gel las librerías estaban listas para ser secuenciadas (Figura 14) mediante la tecnología *Illumina* la cual nos dio la posibilidad de secuenciar conjuntamente en la misma ronda múltiples librerías. Las “lecturas” obtenidas en esta secuenciación pueden ser separadas por librerías de acuerdo a la secuencia de las etiquetas que fueron colocadas en el partidador p7. Los 5 promotores de nuestro interés fueron secuenciados conjuntamente.

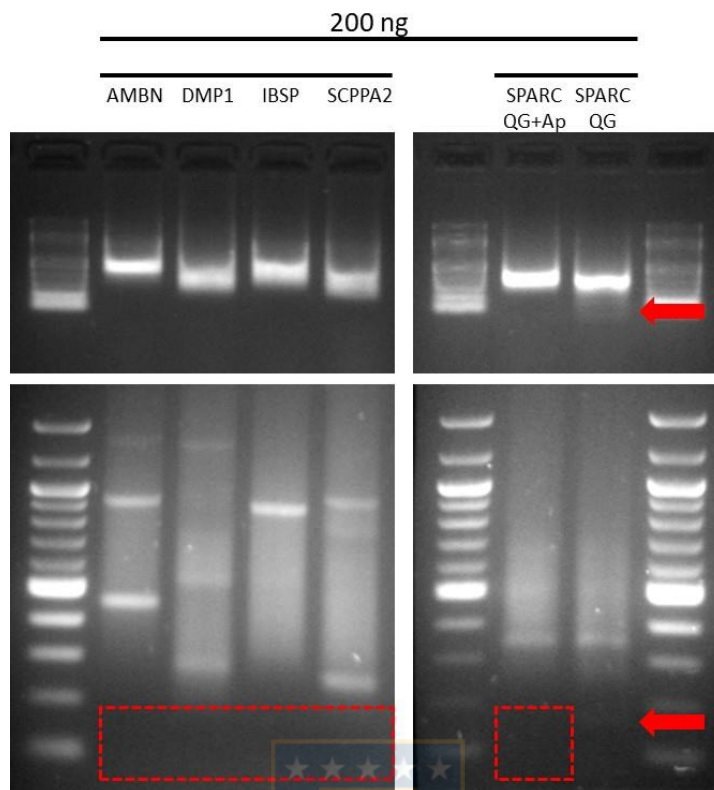


Figura 15: Purificación de librerías por el sistema AMPure

Después de purificar por Qiagen (QG), las librerías fueron re-purificadas con el sistema de beads AMPure (Ap) para asegurar la eliminación de los posibles dímeros de *primers*. En esta figura, sólo el bait de *Sparc* purificado por QG fue comparado con el resultado de purificación de AMPure. Observe que hay un *smear* detectado bajo a banda principal de *Sparc* purificado por QG y desaparece con la re-purificación. Si se compara con la figura 14 se puede observar que las bandas detectadas por debajo de 150 pb y que pueden corresponder a dímeros de *primers* han desaparecido después de la purificación por QG y Ap. Para todas las librerías los valores de absorbancia A260/280 fueron mayores a 1.8 y 260/230 mayores a 2. Todo lo anterior asegura librerías óptimas para secuenciación.

5.6.12 Procesamiento de datos

El ensayo de 4C produce moléculas híbridas de ADN, es decir, una combinación de específicos *vp* (delimitados por partidores) más las secuencias derivadas de los fragmentos de interacción ligados. Debido a que estas secuencias compuestas son imposibles de mapear, fueron recortadas (trimmed) *in silico* para remover la secuencia del *vp*, dejando únicamente el fragmento “capturado” para ser mapeado. Las *lecturas* en formato fastq.gz fueron recortadas usando *cutadapt* (version 2.6). Después de este paso se realizó el mapeo contra el genoma de referencia,

para este trabajo se seleccionó la versión 9.1 de *Xt* descargada en el sitio <http://hgdownload.soe.ucsc.edu/goldenPath/xenTro9/bigZips/> que a pesar de no ser la última versión, está apoyada con datos experimentales y ha sido frecuentemente citada en los últimos años. El alineamiento se llevó a cabo usando el software *Bowtie2* (version 2.3.4.3), el cuál generó archivos de *reads* mapeados (*Sam files*) que fueron transformados a archivos BAM, clasificados e indexados usando las herramientas del paquete *samtools* (version 1.9).

El análisis bioinformático se realizó con el paquete *r3Cseq* diseñado en R por (Supat Thongjuea, 2015) como parte del repositorio de Bioconductor. Debido a que este paquete fue desarrollado para trabajar datos obtenidos en mamíferos (humano, ratón y rata) se diseñaron dos nuevos paquetes que permitieron cargar los archivos de *Xt* (*BSgenome.Xtropicalis.UCSC.xenoTro9* y *BSgenome.Xtropicalis.UCSC.xenoTro9.masked*). De esta forma fue posible importar los archivos Sorted Bam a *r3Cseq*, el cuál generó diferentes archivos (*bedGraph*). Finalmente, estos archivos fueron convertidos a un formato *bigWig* usando la herramienta *bedGraphToBigWig* del sitio UCSC, un formato final binario que permitió visualizar los archivos en el *UCSC genome browser*.

Los datos de secuenciación fueron cargados en el paquete *r3Cseq*. Las lecturas fueron contadas por fragmento de restricción usando la función *getReadCountPerRestrictionFragment*. Con esta función se contaron todas las lecturas dentro de los fragmentos de restricción generados por la enzima *MboI*. El protocolo de 4C siempre genera una muy fuerte señal de interacción en el *vp* y en los fragmentos inmediatamente adyacentes a él (usualmente uno o dos fragmentos). Esta señal es causada ya sea por una falla para digerir el sitio de restricción primario (lo que resulta en la captura de fragmentos vecinos) o por una auto-ligación del fragmento *vp* en si mismo. Estas señales, aunque deberían estar siempre presentes, no son informativas. *r3Cseq* ofrece la función "*nFragmentExcludedReadsNearViewpoint*", la cuál es diseñada para excluir las lecturas alineadas de los fragmentos localizados río arriba y abajo del *vp* (default = 2).

Posteriormente se llevó a cabo la normalización de las lecturas mediante las funciones *calculateRPM* y *calculateBatchRPM* lo que permitió calcular las lecturas por millón por fragmento de restricción (RPM) como valores de frecuencia de interacción normalizada usando el número de lecturas observadas en un particular fragmento de restricción dividido por el número total de lecturas alineadas, multiplicadas por 1 millón. Después de la normalización la función *getInteractions* permite realizar un análisis que asigna valores estadísticos a las regiones de interacción candidatas en *cis* (mismo cromosoma) y *trans* (cromosomas diferentes).

Para identificar las regiones de interacción significativas en *cis* se aplicó un método de corrección de fondo de interacciones. Esto permitió discriminar entre interacciones que son una

simple consecuencia de cortas distancias genómicas al *vp* de aquellas que representan una interacción significativa. Se determinó la relación de las señales 4C-*seq* de regiones genómicas situadas en el cromosoma en *cis* clasificando el recuento de lecturas por región en función de la distancia relativa al *vp* (en 4C las señales de interacción gradualmente decrecen con el incremento de la distancia desde el *vp* y sólo pueden esporádicamente ser capturadas en otros cromosomas). A continuación, se aplicó un algoritmo *spline* de suavizado cúbico de regresión no paramétrica implementado en R con el parámetro de suavizado que por default es 0.1 (Figura 15). Se asumió que una fracción relativamente pequeña de interacciones podría significativamente interactuar con los fragmentos (promotores) de estudio, por lo tanto, se utilizaron las señales de interacción escaladas promedio como la señal esperada para una distancia genómica determinada. Las señales 4C-*seq* en *cis* fueron entonces transformadas en un Z-score usando la formula $(obs - exp)/SD$, donde *obs* es la señal 4C detectada para una interacción específica, *exp* es el promedio ponderado calculado de las señales 4C para una distancia genómica específica y *SD* es a desviación estándar ponderada calculada de los valores residuales (*obs - exp*). Una vez que los valores de Z fueron calculados, los valores de P fueron entonces asignados a cada Z-score y transformados en un q-value para un análisis FDR usando el paquete q-value de Bioconductor con un nivel FDR de 0.05. [qvalue(interactions.p.values, fdr.level = 0.05, pi0.method = 'bootstrap')].

Para evaluar las interacciones en *trans* se aplicó una formula similar a la anterior asumiendo que las interacciones capturadas en *trans* podrían tener una señal de interacción más alta que los valores de interacción global. Por lo tanto, se transformaron la señales de interacción detectada en un Z-score con la formula “ $(obs-exp)/SD$ ”, donde *obs* fue la señal de interacción observada encontrada en todo el conjunto de datos (excluyendo las regiones localizadas dentro de ± 100 kb alrededor del *vp*), *exp* fue la señal de interacción promedio para todo el conjunto de datos y *SD* es la desviación estándar del conjunto total. A continuación, un procedimiento similar al realizado para las interacciones en *cis* fue realizado para transformar la señal de interacción en *trans* en valores de interacción estadística (*P* y q-values)

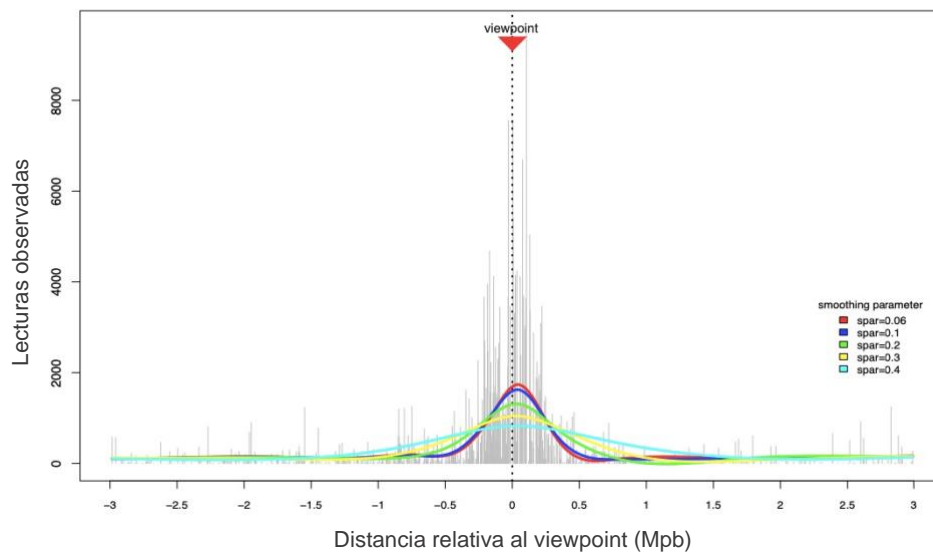


Figura 16: **Normalización de las señales de interacción**

Escalado de las señales de interacción de fondo utilizando un algoritmo de spline de suavizado cúbico de regresión no paramétrica en R. Las señales de interacción se clasificaron en función de la distancia genómica relativa al *vp*. El gráfico muestra las señales de interacción observadas (gris) y la señal de interacción escalada utilizando diferentes parámetros de suavizado dentro de una ventana de ± 3 Mpb alrededor del *vp* (imagen modificada de (Thongjuea, 2015)).

Los métodos estadísticos propuestos para este trabajo están ejemplificados en el sitio <http://r3cseq.genereg.net/Site/index.html> y han sido aplicados para identificar interacciones en *cis* exitosamente en análisis de datos de 4C y 5C (Thongjuea et al., 2013). Una explicación detallada de los procesos de normalización y análisis de interacciones en *cis* y *trans* está bien descrita en el *pipeline* de (Thongjuea, 2015).

6. RESULTADOS

6.1 Expresión espacial específica de los miembros del locus SCPP y SPARC

Dos estudios realizados en nuestro laboratorio dieron pauta para este trabajo, en uno de ellos se demostró que la expresión de los genes *Ibsp*, *Dmp1* y *Sparc* en hueso de *Xt* reflejan una alta conservación con mamíferos, así como un patrón de expresión divergente entre osteoblastos y osteocitos. *Dmp1* se detectó fuertemente en osteocitos, pero no en osteoblastos. En cambio, *Ibsp* y *Sparc* fueron detectados fuertemente en osteoblastos, pero sólo débilmente en algunos osteocitos, principalmente en las células recientemente diferenciadas (es decir, las más cercanas a la superficie ósea), por lo que se concluyó que los osteocitos son molecularmente heterogéneos (Espinoza et al., 2010). En otro estudio, llamó la atención que la expresión de los genes *Col1a1* y *Col1a2* (Enault et al., 2015) fueran decaendo desde los osteocitos más superficiales hasta los más profundos, es decir de los más jóvenes a los más antiguos (Figura 1A). Estas observaciones podrían deberse a que enhancers osteocíticos van perdiendo su actividad transcripcional a medida que dichas células “envejecen” (es decir, a medida que los osteocitos pasan más tiempo embebidos en la matriz). Otra posibilidad es que dichos genes están controlados exclusivamente por enhancers osteoblásticos, y que la señal decreciente observada refleje la lenta degradación de los transcritos en el linaje osteocítico. Al respecto, se sabe que los ARNm pueden tener una vida media larga y permanecer en células en las cuales el gen desde que fue expresado ya se apagó transcripcionalmente (Figura 1B). Por lo tanto, no hay una correlación exacta entre la presencia de transcritos maduros y el estado transcripcional del gen. Por este motivo, los inestables ARN intrónicos indican de mejor forma en cuál célula exactamente un gen dado está siendo activamente transcrito en el contexto de un embrión en desarrollo (Dubrulle & Pourquié, 2004). Para discriminar entre ambas posibilidades realizamos hibridaciones *in situ* con sondas intrónicas. Una sonda intrónica permitirá detectar la señal exclusivamente en núcleos de osteoblastos si se diseña contra un gen que está siendo expresado exclusivamente en este tipo celular, y que se apaga transcripcionalmente durante la diferenciación a osteocitos (Figura 17C). La detección con sonda intrónica sólo es posible en el núcleo celular (antes de la maduración del ARN mensajero), por lo tanto, como control interno se incluyó el marcador nuclear DAPI para asegurarse que la marca obtenida co-localiza con este organelo celular.

Debido a que se sabe que la familia de genes *SCPP* ácidos tiene 5 miembros en *Xenopus*, se utilizaron los datos de RNA-seq de nuestro laboratorio para corroborar la expresión

de los miembros SCPP en hueso. Usando estos datos, identificamos también la expresión de *Scppa2*, un gen que no había sido descrito en nuestro laboratorio pero que había sido observado fuertemente expresado en osteoblastos y osteocitos en hueso mandibular (hueso maxilar superior) de *Xt* (Kawasaki, 2009). Las características de *Scppa2*, tanto respecto a su ubicación en el genoma como a su función y su patrón de expresión espacial, se asemejan al ortólogo *Spp1* de mamíferos.

Para realizar las hibridaciones *in situ* se seccionan longitudinalmente las extremidades posteriores de renacuajos en el estadio NF60 en la región tibiofibular (los huesos tibia y fíbula se fusionan en un solo hueso durante el desarrollo de *Xt* (Figura 1D). Se seleccionan los mejores cortes con base en características histológicas mediante tinción con hematoxilina-eosina. En este estadio el hueso endocondral está en formación y al seccionarse pueden distinguirse músculo estriado, cartílago, médula y hueso (Figura 1E). Rodeando la matriz ósea calcificada se encuentra una capa de tejido conectivo que recubre al hueso denominada periostio y desde el interior, también rodeando la matriz ósea, el endostio. En ambas regiones se aprecia una fina capa de osteoblastos que en la imagen de la figura 1E es evidente en el periostio. El rectángulo rojo de la figura 1E es ampliado en 1F, donde es posible observar cómo al interior de la matriz ósea se encuentran embebidos los osteocitos, células que derivan de los osteoblastos. A nivel de la diáfisis hay formación de médula ósea con abundantes vasos sanguíneos y cartílago hecho de condrocitos que será reemplazado por hueso trabecular. Es importante destacar una zona menos densa denominada osteoide que es secretada por los osteoblastos y que se caracteriza por ser matriz no calcificada. También puede apreciarse la matriz calcificada secretada por los condrocitos que están en contacto con el endostio (Figura 1F).

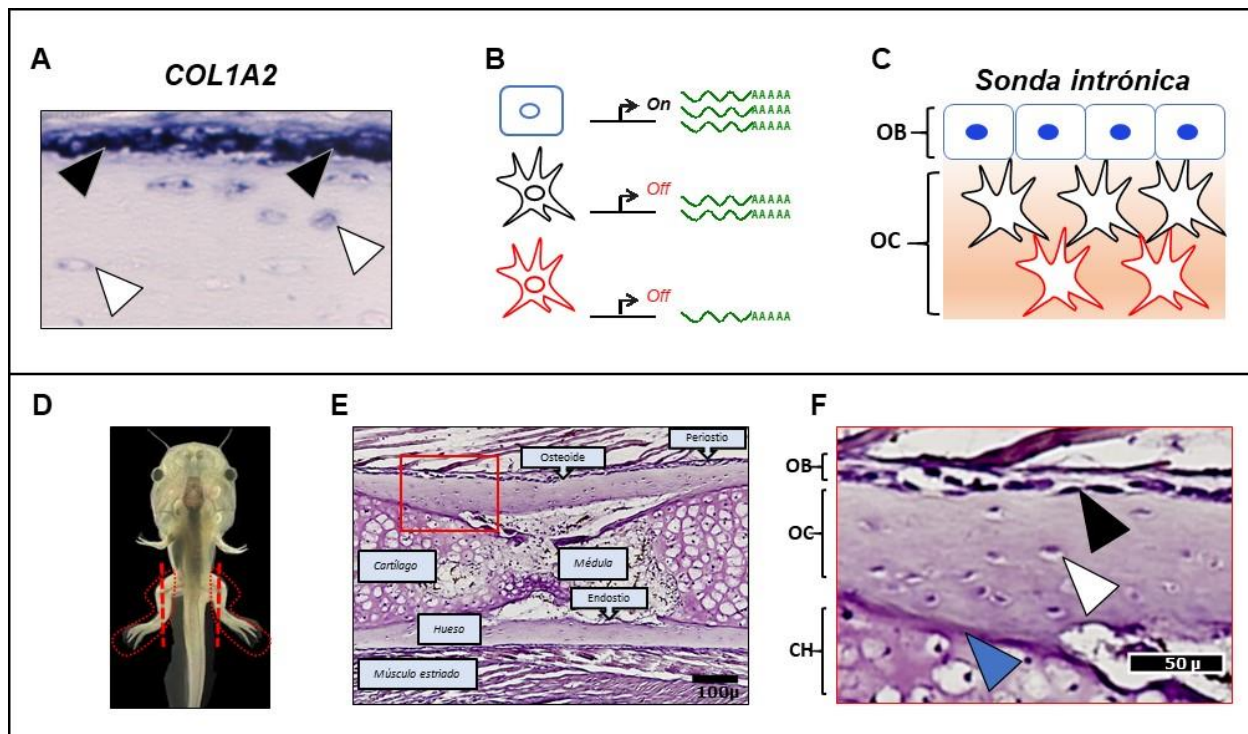


Figura 17: Antecedente experimental e histología del hueso

En **A** se muestra la detección de ARNm por hibridación *in situ* del gen *Col1a1* en una sección de fémur de *Xt NF60*. La capa celular osteoblástica (OST) muestra una fuerte señal (puntas de flecha negras) mientras que la señal es débil en los osteocitos (OC) que están embebidos en la matriz del hueso (Enault, 2015). Se observa que la expresión del *Col1a1* en osteocitos va disminuyendo gradualmente desde los recién-diferenciados hasta los más viejos donde la detección se vuelve cada vez más tenue (puntas de flecha blancas). En **B**, un esquema muestra que un gen puede estar transcripcionalmente activo (*On*) con abundantes transcritos, sin embargo, en el modelo que proponemos, algunos transcritos podrían permanecer estables en el citoplasma celular de osteocitos que diferencian directamente de osteoblastos aun cuando el gen ya está apagado (*Off*). La disminución de transcritos es evidente en osteocitos más profundos o más “viejos” (contorno rojo). En **C**, si lo anterior es correcto, una sonda intrónica, permitiría detectar los ARN mensajeros sólo en núcleos de osteoblastos, pero no en osteocitos. En cambio, *Dmp1* sería detectado sólo en núcleos de osteocitos (no mostrado). En **D**, el modelo de estudio *Xt* en estadio de renacuajo, justo antes de que se lleve a cabo la metamorfosis a rana juvenil. Las extremidades posteriores son embebidas en parafina y seccionadas longitudinalmente para estudios de histología con HES. En **E**, una sección de hueso tibiofibular muestra los tejidos de una sección de 7 µm. Pueden distinguirse fácilmente (de afuera hacia adentro) las fibras del músculo estriado, el periostio, la fina capa de osteoblastos, el hueso, una fina capa de tejido conectivo o endostio, y el cartílago, junto a la médula ósea en formación en una región de abundantes vasos sanguíneos. El recuadro rojo es ampliado en **F**, donde se aprecia mejor la fina capa de osteoblastos (punta de flecha negra) que se diferencian a osteocitos (punta de flecha blanca), así como la fina capa de tejido denso del endostio, que contiene algunos osteoblastos y que está en contacto con condrocitos que secretan la matriz calcificada (punta de flecha azul).

6.1.1 *DMP1* se expresa específicamente en osteocitos

Los transcritos maduros de *Dmp1* se detectaron con una sonda exónica fuertemente en citoplasma de osteocitos (Figura 2 A y C). Además, la sonda intrónica, que hibrida con los mensajeros recién sintetizados de *Dmp1*, permitió corroborar la señal exclusivamente en el núcleo de osteocitos (Figura 2 E y G). No se detectó expresión en osteoblastos, ni con la sonda exónica ni con la sonda intrónica.

Para asegurarse que la marca detectada con sondas intrónicas es exclusivamente nuclear y no se trata de un artefacto derivado de la reacción enzimática de la fosfatasa alcalina, se realizó una tinción de los núcleos con DAPI y con el software *ImageJ* se llevó a cabo una superposición de imágenes de campo claro y fluorescente (Figura 2, columna derecha). En la superposición se invirtió el campo claro y se asignó un color rojo a la marca de in situ, lo que generó imágenes de pseudocolor en las que una marca rosa implica una superposición de la marca de in situ en color rojo con el azul de los núcleos mientras que un color amarillo es derivado del ruido de los elementos del tejido, por ejemplo, de vasos sanguíneos y células pigmentadas (esquema explicativo en la Figura 2J).

Gracias a la superposición es posible apreciar la diversidad celular del tejido, por ejemplo, la fina línea de osteoblastos que se encuentra en el periostio muestra como células que tan pronto quedan atrapadas en su matriz, expresan fuertemente *Dmp1* (Figura 2 B y D). A mayor aumento, se observa que la marca detectada con sondas intrónicas es exclusivamente nuclear (H). Mediante la superposición también fue posible observar que algunos osteocitos son transcripcionalmente inactivos, esto es evidente en las imágenes B y D, donde en un mismo corte es posible apreciar células cuya marca de in situ es tenue o nula en citoplasma (Figura D). En resumen, estas observaciones permitieron sugerir que el gen *Dmp1* se activa tan pronto los osteoblastos maduros diferencian a osteocitos.

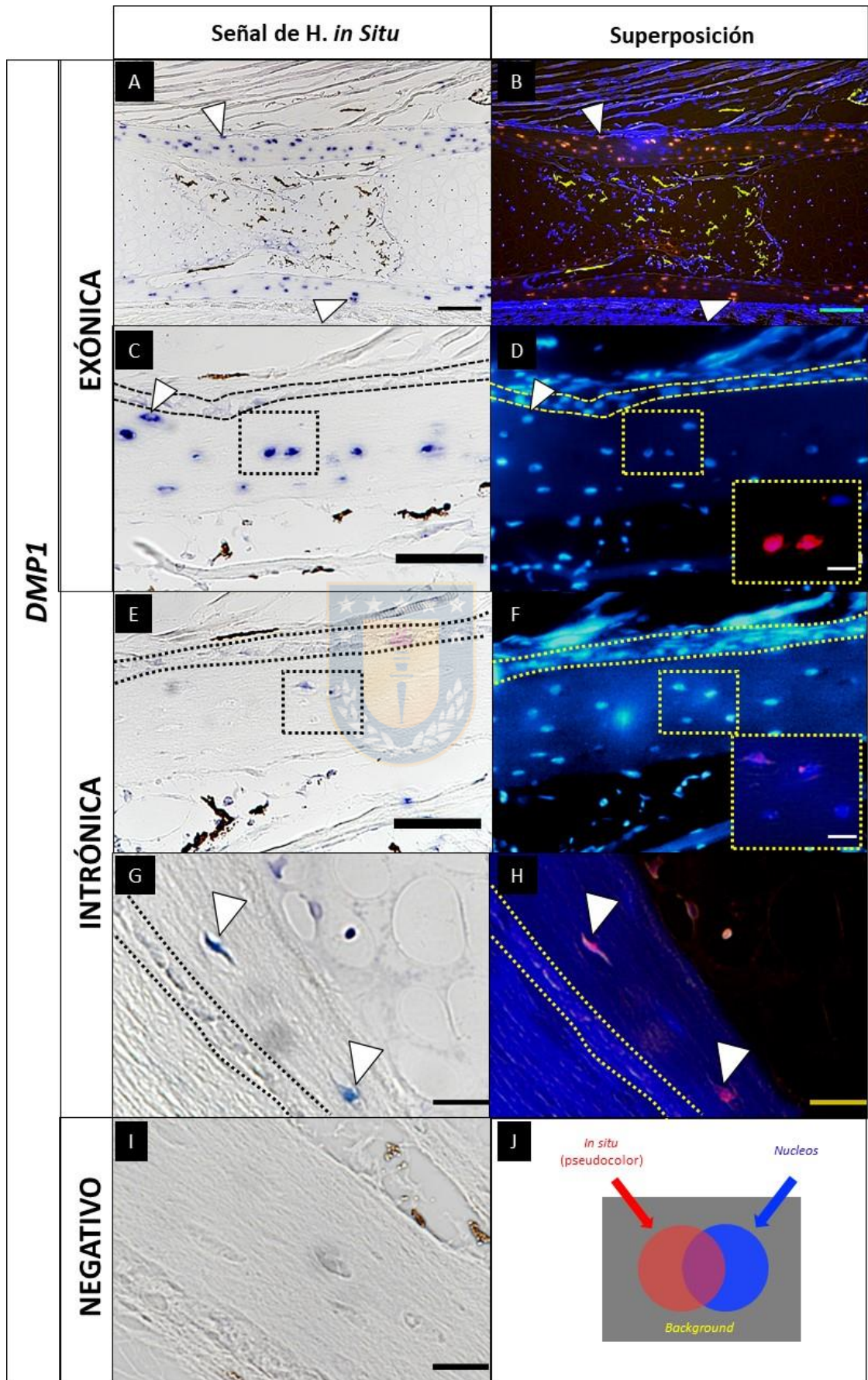


Figura 18: Expresión espacial de *Dmp1* con sonda exónica e intrónica

En la columna de la izquierda se observa la señal de hibridación *in situ* detectada por microscopia de campo claro en secciones de hueso tibiofibular (7 μ). En la columna derecha se observan los núcleos teñidos con DAPI (azul fluorescente) en superposición con la imagen de campo claro invertida. En A y B se muestra que *Dmp1* se activa tan pronto las células quedan atrapadas en su matriz (puntas de flechas blancas). Un aumento en C y D permite apreciar mejor la fuerte expresión de *Dmp1* sólo en osteocitos. Dos líneas punteadas (C-H) permiten delimitar la capa de osteoblastos del hueso mineralizado donde se encuentran embebidos los osteocitos. Usando sondas intrónicas se confirmó la expresión de *Dmp1* sólo en osteocitos (E-H), un aumento permite observar la superposición de la marca nuclear en osteocitos (rectángulo amarillo en F y flechas blancas en G y H). Los núcleos al superponerse con la marca de *in situ* invertida se visualizan en un tono rosado (H). En J, los colores superpuestos en rojo (pseudocolor de señal *in situ*) y azul (DAPI) en un fondo gris (imagen de campo claro invertida) permiten evaluar si la señal detectada por hibridación *in situ* es exclusivamente nuclear. La señal de ruido, propia del tejido se detecta en un color amarillo (por ejemplo, los vasos sanguíneos). Para el control negativo se utilizaron las sondas "sentido" exónica del gen *Dmp1*. Se realizaron al menos 3 réplicas experimentales. Tamaño de barras: 100 μ m (A y B); 50 μ m (C y E); 20 μ m (G, H, I); 10 μ m (magnificación en D y F).

6.1.2 *IBSP* se expresa específicamente en osteoblastos

Los transcritos de *lbsp* fueron detectados con sondas exónicas en citoplasma de osteoblastos y en algunos osteocitos. La señal exónica fue detectada fuertemente en osteoblastos en el endostio a nivel de la diáfisis (Figura 2 A-D), no obstante, con la sonda intrónica se detectó señal nuclear exclusivamente en osteoblastos, y no en osteocitos. (Figura 3 E y G). Esto sugiere que la señal citoplasmática (exónica) es el producto de los transcritos estables tardíos de *lbsp*.

Del punto de vista regulatorio, nuestros resultados apoyan la idea que en *Xt* los MCRs que controlan la expresión de *lbsp* están activos sólo en osteoblastos y la señal débil observada en algunos osteocitos puede ser el ARN mensajero maduro tardío. Considerando que osteocitos diferencian directamente de osteoblastos, la expresión de *lbsp* revela un dramático switch regulatorio en la transición osteoblasto-osteocito. *lbsp* es abruptamente apagado mientras *Dmp1* es rápidamente activado (Figura 4).

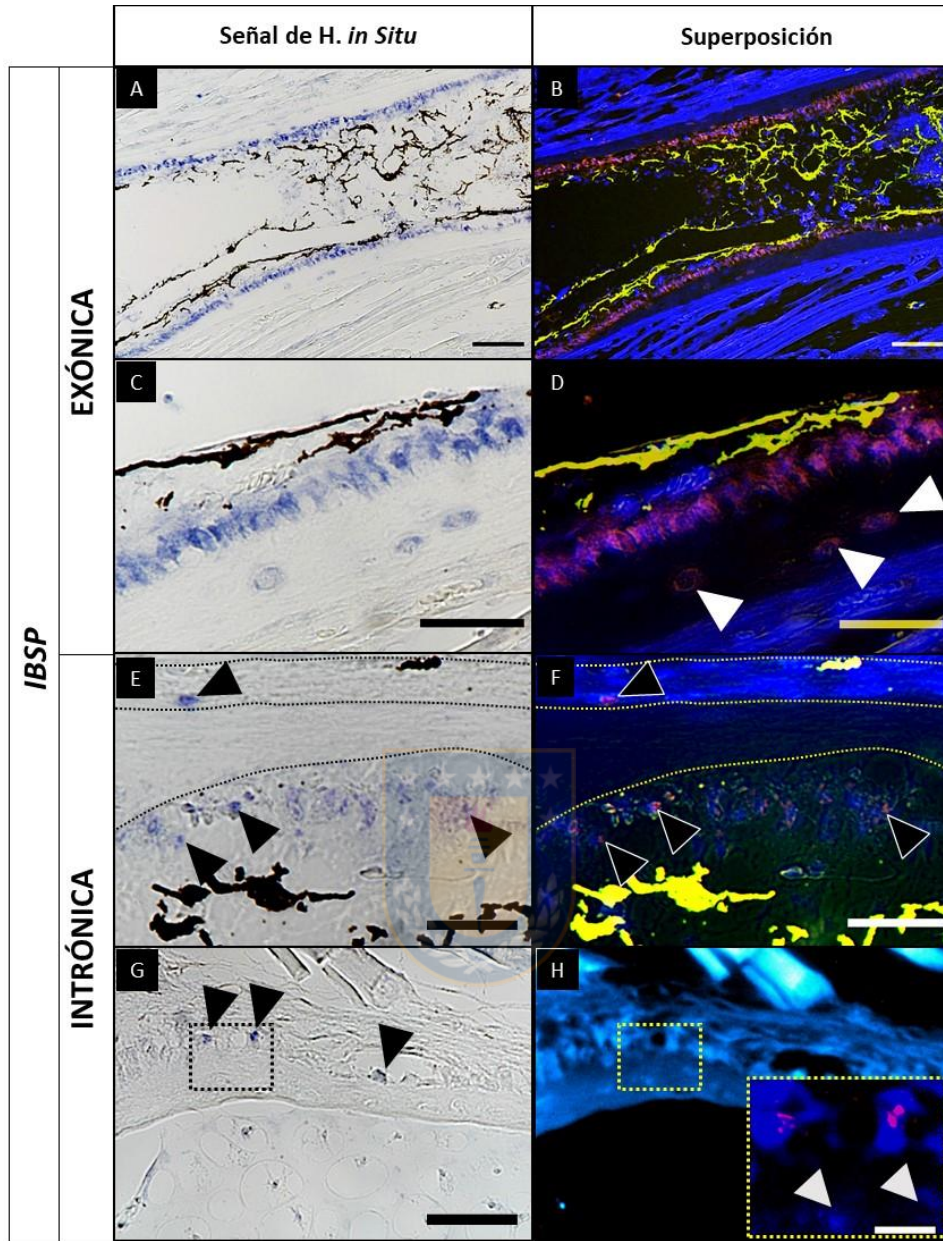


Figura 19: **Expresión espacial de *lbsp* con sonda exónica e intrónica**

Se detectaron transcritos maduros (sondas exónicas) fuertemente en citoplasma de osteoblastos. En esta imagen la expresión de *lbsp* es fuerte en los osteoblastos del endostio (A-D). Un aumento en C y D permite apreciar mejor la fuerte expresión de *lbsp* sólo en osteoblastos y débilmente en algunos osteocitos (flechas blancas). En E, las sondas intrónicas hibridaron con los ARN mensajeros inmaduros específicamente en osteoblastos de la región del endostio. La superposición de imágenes muestra que la marca es exclusivamente nuclear (flechas negras). Los puntos o regiones amarillas en la superposición son vasos sanguíneos que están enriquecidos en la médula ósea. En G se muestran algunos osteoblastos que fueron detectados específicamente a nivel del periostio cuya marca nuclear es evidente en H (recuadro amarillo). Se realizaron al menos 3 réplicas experimentales. Tamaño de barras: 100 μ m (A y B); 50 μ m (C-G); 10 μ m (magnificación en H).

6.1.3 *Sparc* se expresa específicamente en osteoblastos

Sparc fue específicamente detectado con sondas exónicas en la capa de osteoblastos del periostio (Figura 5 A) así como en regiones activas de osteoblastos (flechas blancas en Figura 5A). En todas las secciones analizadas para este gen fue posible detectar una señal débil en osteocitos (Figura B) en algunas regiones del periostio. Esta señal va decayendo gradualmente a medida que los osteoblastos se diferencian a osteocitos de forma similar al patrón observado con *Ibsp*. La sonda intrónica mostró señal únicamente en osteoblastos, lo que lleva a sugerir que, al igual que con *Ibsp*, la actividad de este gen es exclusiva de osteoblastos. Por lo tanto, los transcritos remanentes detectados con sondas exónicas sugieren una vida media larga de los mensajeros en osteocitos detectable, aún cuando el gen *Sparc* ya está apagado. Un acontecimiento molecular similar al que se propone para *Ibsp* en la figura 4.

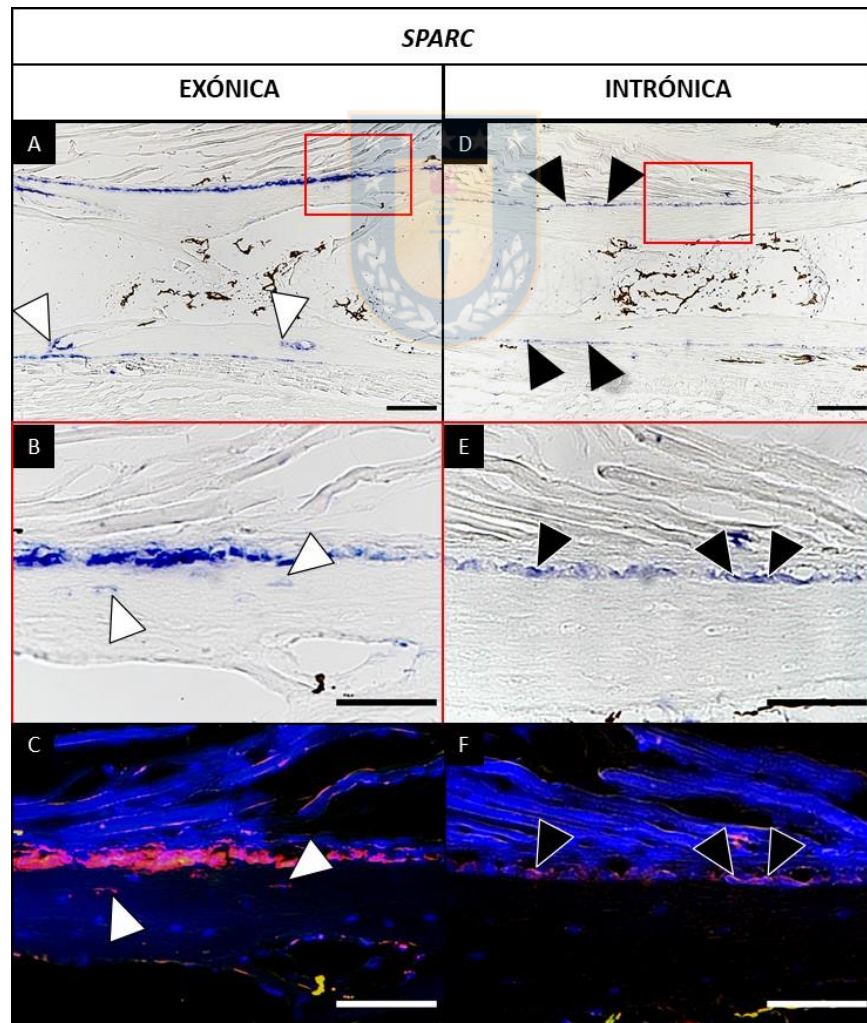


Figura 20: Expresión espacial de *Sparc* con sonda exónica e intrónica

En **A** se observa la detección de la señal de hibridación *in situ* con sonda exónica exclusivamente en osteoblastos de la capa del periostio, así como en algunas regiones donde los osteoblastos están activos (flechas blancas). Un aumento de la figura **A** (recuadro rojo) muestra que la señal exónica puede ser también detectada en el citoplasma de algunos osteocitos próximos a la capa de osteoblastos recién diferenciados (**B**). Observe cómo en estos osteocitos (**B**), la señal va decayendo gradualmente de forma similar a la expresión del gen de *Col1a2* que se presentó en la figura **1A**. Aunque la señal citoplasmática en osteocitos es débil, aún se puede apreciar en la imagen de superposición en **E** (puntas de flecha blancas). En la columna de la derecha se compara la expresión del gen *Sparc* mediante sonda intrónica. En la figura **D** se aprecia que la señal intrónica sólo es detectable en la capa de osteoblastos del periostio (puntas de flecha negra). Un aumento de la región del recuadro rojo en **E** permite apreciar de mejor forma la señal detectable sólo en osteoblastos y no en osteocitos. La imagen de superposición en **F** permite corroborar que la señal observada en **E** es exclusivamente nuclear. La detección con sonda intrónica y exónica de *Sparc* se replicó dos veces. Tamaño de barras: 100 µm (A y B) y 50 µm (B, C, E y F).

Con todo lo anterior se propone un modelo de la **transición de osteoblasto a osteocito** (Figura 20A). Los osteoblastos producen el osteoide o matriz no mineralizada. Genes como *Ibsp* y *Sparc* son activos (*On*) en osteoblastos. Con el tiempo, los osteoblastos continúan secretando la matriz y quedan atrapados en ella en el momento que se diferencian a osteocitos. De esta forma, la señal de *in situ* obtenida con una sonda exónica detecta fuertemente transcritos de *Ibsp* y *Sparc* en osteoblastos y débilmente en el citoplasma de osteocitos, incluso cuando ambos genes han sido apagados transcripcionalmente en osteocitos (*Off*) debido probablemente a los transcritos tardíos que permanecieron en las células recién-diferenciadas. Por su parte, *Dmp1* es inactivo en osteoblastos, pero dramáticamente sobrerregulado en osteocitos, donde puede ser detectado fuertemente con sondas exónicas (señal citoplasmática) e intrónicas (señal nuclear). La señal en osteocitos con la sonda intrónica de *Dmp1* no es visible en todas las células porque el ARNm es rápidamente procesado, dificultando la visualización de una señal intrónica. Sin embargo, aún así fue posible observar la expresión de *Dmp1* de forma específica en el núcleo de osteocitos gracias a una tinción nuclear con DAPI.

Nuestros resultados revelan un dramático *switch* regulatorio en la transición osteoblasto osteocito. Respecto al locus SCPP, en el tiempo en que *Ibsp* es abruptamente apagado, *Dmp1* es rápidamente activado. En la figura 20B, se muestra la región del locus SCPP que contiene los genes *Dmp1* e *Ibsp*. Debido a que *Dmp1* e *Ibsp* tienen un patrón de expresión diferencial aun cuando tienen un orden sinténico en el genoma es posible sugerir que hay MCRs específicos de osteocitos que activan *DMP1* y específicos de osteoblastos que activan a *IBSP*.

Para comprobar lo anterior fue necesario primero demostrar que existen MCRs que activan los genes del locus SCPP y de SPARC en osteoblastos de cultivo primario, por lo que se recurrió a estudiar la funcionalidad de dichos elementos putativos.

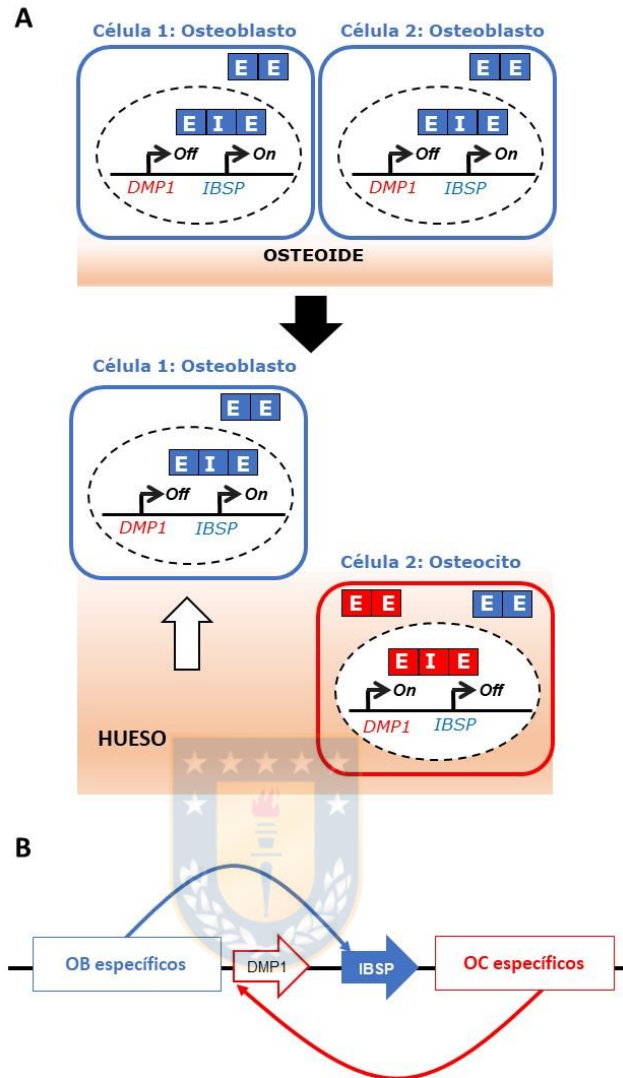


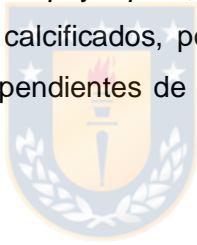
Figura 21: Fuerte **switch** regulatorio en la transición osteoblasto-osteocito

Los osteoblastos (células de contorno azul) producen el osteoide o matriz no mineralizada. En este esquema, *lbsp* es activo (*On*). Los osteoblastos continúan secretando la matriz y quedan atrapados en el momento que se diferencian a osteocitos (célula 2 de contorno rojo). En este modelo se explica porqué la señal de in situ con una sonda exónica detecta fuertemente transcritos de *lbsp* en osteoblastos y una débil señal en el citoplasma de osteocitos, incluso cuando el gen *lbsp* ha sido apagado transcripcionalmente en osteocitos (*Off*). Por su parte *Dmp1* es inactivo en osteoblastos, pero dramáticamente sobrerregulado en osteocitos, donde puede ser detectado tanto con sondas exónicas (señal citoplasmática) como intrónicas (señal nuclear). Nuestros resultados revelan un dramático **switch** regulatorio en la transición osteoblasto osteocito. En el tiempo en que *lbsp* es abruptamente apagado, *Dmp1* es rápidamente activado. En B, se muestra la región del locus SCPP que contiene los genes *Dmp1* e *lbsp*. Debido a que estos genes tienen un patrón de expresión diferencial, aun cuando tienen un orden sinténico en el genoma, es posible hipotetizar que hay MCRs específicos de osteocitos que activan *Dmp1* y específicos de osteoblastos que activan a *lbsp*.

6.1.4 SCPPA2 tiene un patrón de expresión divergente

En todas las secciones de tejido óseo tibiofibular de *Xt* el gen *Scppa2* se expresó sólo en algunas regiones de la capa de osteoblastos, así como en algunos osteocitos, los cuáles, a diferencia de *Ibsp* o *Sparc*, presentaron una fuerte señal citoplasmática (Figura 6 A y B) de forma similar a las observaciones previas hechas por (Kawasaki, 2009) en hueso mandibular. Sin embargo, el hecho de haber seleccionado el hueso de extremidad, el cuál deriva de un proceso de osificación endocondral, nos permitió observar que *Scppa2* se expresa también fuertemente en cartílago calcificado, el cuál consiste en un área de algunas células de espesor que están en contacto con el endostio a nivel de la diáfisis. El análisis con sonda intrónica muestra expresión de *Scppa2* en osteoblastos y en algunos condrocitos del cartílago calcificado, sin embargo, en estas secciones no fue posible tener una buena tinción con DAPI, lo que imposibilitó realizar una imagen de superposición. No obstante, en C y D se observan algunos osteoblastos (flechas negras) con una señal tenue, la cuál también se pudo detectar parcialmente en núcleos de osteocitos en C.

En conclusión, a diferencia de *Ibsp* y *Sparc*, *Scppa2* se expresa espacialmente en osteoblastos, osteocitos y condrocitos calcificados, por lo que probablemente su regulación transcripcional es debido a MCRs independientes de los que controlan la expresión de otros genes del locus *Scppa2*.



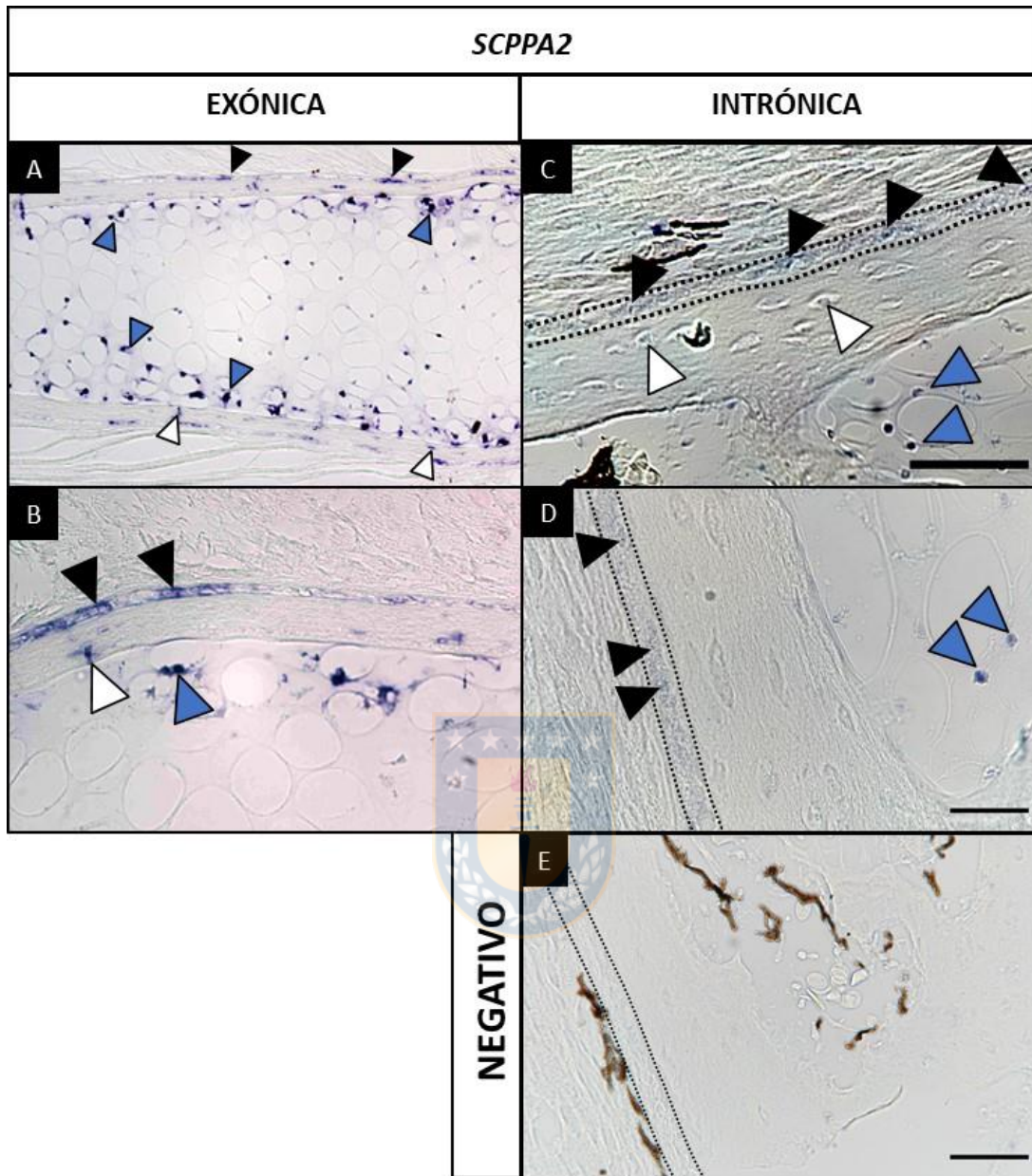


Figura 22: **Patrones de expresión de *Scppa2* con sonda exónica e intrónica**

En A se muestra la vista global de la expresión de *Scppa2* en extremidad hueso fibiotibial de *Xt* en estadio NF60. Los osteoblastos, osteocitos y condrocitos son señalados con puntas de flecha de color negro, blanco y azul respectivamente. En B, se aprecia mejor la fuerte expresión en algunas secciones de la capa de osteoblastos, así como en algunos osteocitos y fuertemente en condrocitos calcificados. Un patrón de expresión similar fue observado también en vértebras del mismo estadio (datos no mostrados). En C y D, con sonda intrónica, fue posible detectar *Scppa2* en algunos núcleos de osteoblastos (flechas negras y en condrocitos (flechas azules). Debido a que la señal es poco perceptible en osteocitos (flechas blancas en C), no es posible concluir la expresión de *Scppa2*. Para el negativo se usaron ambas sondas, exónica e intrónica del gen *Scppa2*. Sólo se muestra la sonda intrónica. Tamaño de barras: 50 μ en C y 20 μ en D y E.

6.2 Análisis funcional de los MCRs del locus SCPP y de SPARC

En este trabajo fue fundamental demostrar funcionalmente que regiones previamente identificadas por ChIP-seq en nuestro laboratorio y que corresponden a enhancers putativos permiten explicar los patrones de expresión de los genes SCPP y SPARC descritos en la sección anterior. Sabiendo que las marcas H3K4me1 y H3K27ac se asocian frecuentemente con MCR activos se buscaron marcas cercanas al locus SCPP y al gen SPARC tanto en regiones intergénicas como intrónicas. Se seleccionaron las regiones que contenían los MCRs mejor enriquecidos con ambas marcas de histonas, se amplificaron y clonaron en plásmidos de expresión para ser probados funcionalmente en ensayos reporteros-GFPn en osteoblastos de cultivo primario. Los plásmidos fueron electroporados y el resultado de fluorescencia de cada célula fue cuantificado mediante citometría de flujo.

La expresión de GFPn, cuantificada en función de la razón GFP/mCherry en osteoblastos, reflejó la fuerza transcripcional del enhancer. Se probaron 6 enhancers putativos del gen SPARC y 4 enhancers del locus SCPP con sus respectivos controles negativos. Estos últimos correspondieron a los promotores basales de cada gen de interés y los mismos promotores acoplados río arriba a una región sin marcas de enhancer, el cuál fue denominado Control Negativo Intrónico (CNI). En general, para ensayos reporteros, se utiliza un promotor basal “universal” para probar la actividad de diversos enhancers. Sin embargo, decidimos trabajar con los promotores de nuestros genes de interés dado que se ha demostrado una gran especificidad entre enhancers y promotor, y que no todos los enhancers son capaces de activar todos los promotores (Gehrig et al., 2009).

6.2.1 Expresión de miembros del locus SCPP y SPARC en hueso

La expresión de los genes SCPP fue confirmada usando los datos de RNA-seq obtenidos en nuestro laboratorio a partir de muestras de hueso de calvaria de Xt(NF60) y mesénquima (NF54). 3 genes del locus SCPP se expresan en hueso y mesénquima (Figura 23). *Dmp1* es fuertemente expresado en hueso y prácticamente nulo en mesénquima mientras que *lbsp* es expresado en ambos tejidos. *Scppa2* se expresa en mesénquima, pero sus niveles son mayores en hueso (Figura 23A). Debido primero, a que el RNA-seq de hueso se realizó en muestras de calvaria, la cuál incluye ambos tipos celulares (osteoblastos y osteocitos) y segundo, que los experimentos de ensayos reporteros se realizaron en osteoblastos de cultivo primario, fue importante confirmar la expresión mediante RT-PCR de los genes SCPP que se expresan específicamente en

osteoblastos primarios (Figura 23B). Con este resultado se confirmó lo observado previamente por hibridación *in situ*, los genes SCPP que se expresan específicamente en osteoblastos son *lbsp* y *Scppa2* (Figura 23B) mientras que *Dmp1* y *Ambn* son prácticamente nulos. Al respecto, se sabe que *Dmp1* sólo se expresa en pasajes tardíos de cultivo primario (día 12), justo en el momento en que las células inician una fase de mineralización (Bertin et al., 2015). En la imagen de la Figura 23B, el gen *Sparc* se expresa también en osteoblastos y *Ef1a*, es un gen ubicuo que sirvió como control positivo.

En el laboratorio también tenemos acceso a los datos de RNA-seq obtenidos de otros tejidos de *Xt* (colaboración con Francia). La expresión de los miembros del locus SCPP es prácticamente nula en otros tejidos (riñón, cerebro, hígado, intestino, músculo) excepto para *Scppa2*, un gen que es altamente expresado en riñón. A diferencia de los miembros del locus SCPP, *Sparc* es altamente expresado en todos los tejidos, demostrando su naturaleza semi-ubicua (datos no mostrados), consistente con un posible rol de chaperona extracelular de colágeno (Martinek et al., 2007).

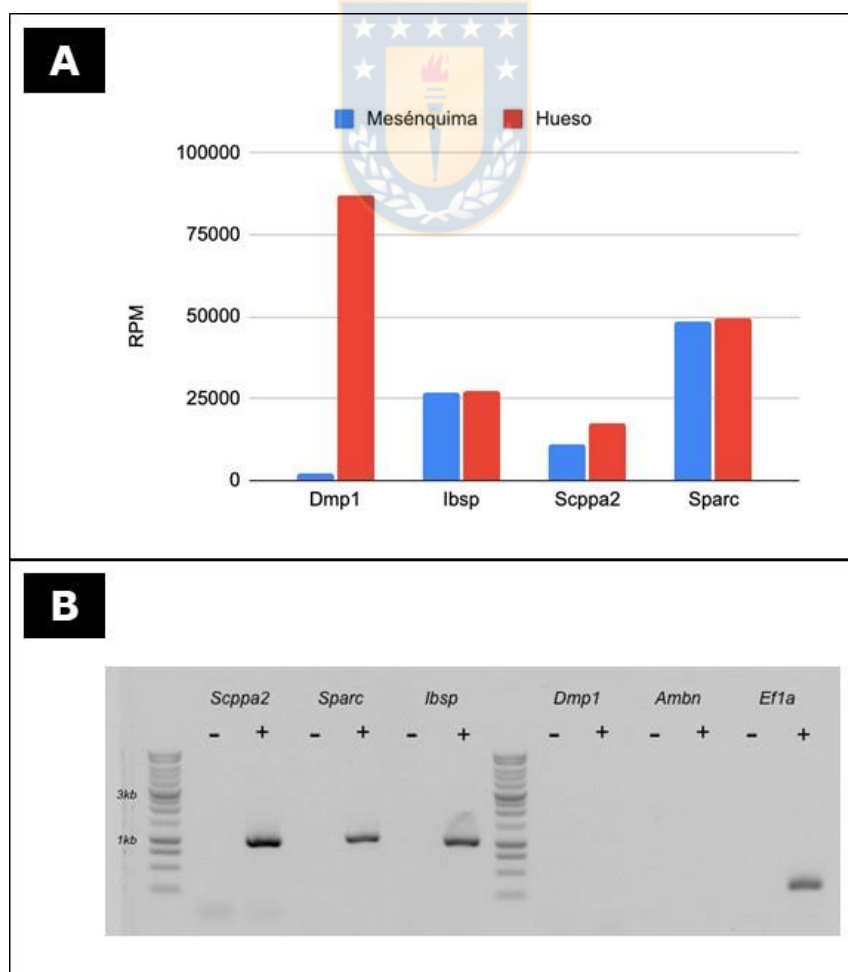


Figura 23: Expresión de genes del locus SCPP y SPARC

En A se muestran los niveles de expresión relativa derivados de experimentos de RNA-seq de los genes SCPP ácidos que se expresan en hueso y/o mesénquima. *Dmp1*, *Ibsp* y *Scppa2* son expresados fuertemente en hueso. En B, se extrajo ARN de osteoblastos de cultivo primario en pasaje 2, se retro-transcribió y el producto fue utilizado como templado para amplificar los genes SCPP y evaluar su expresión de forma semicuantitativa. Se confirmó la expresión de los genes *Ibsp* y *Scppa2* en osteoblastos. El gen *Sparc* también fue amplificado, así como un control positivo interno (*Ef1a*). La expresión de *Dmp1* es prácticamente nula en el pasaje 2 de cultivo primario de osteoblastos, lo cual concuerda con lo publicado por (Bertin et al., 2015), quien mostró que *Dmp1* sólo es expresando en pasajes posteriores o en proceso de mineralización. Como control negativo se utilizó *Ambn*, un gen perteneciente al locus SCPP-P/Q que se expresa principalmente en ameloblastos y no en hueso. La amplificación se realizó utilizando como negativo el templado que fue retro-transcrito sin la enzima *superscript* RT (n=3).

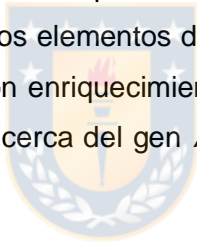
6.2.2 Paisaje regulatorio del locus SCPP

Los datos de RNA-seq cargados en el genoma browser de nuestro laboratorio nos permitieron reconocer espacialmente el inicio (5') y fin (3') de cada gen, así como los exones, intrones y regiones intergénicas. Un alineamiento con el genoma versión 9.1 de *Xt* nos permitió corroborar cada uno de estos elementos genómicos. En la figura 24 se observa una ventana de 150 kbp que abarca el locus SCPP, así como los niveles relativos de expresión de los 3 genes del locus expresados en mesénquima y hueso. Al respecto, es importante considerar que los genes *Scppa1* Y *Mepe* descritos por Kawasaki (2009) en *Xt* y que se encuentran entre *Dmp1/Ibsp* e *Ibsp/Scppa2* respectivamente no se expresan en mesénquima y hueso. A propósito, el gen descrito por (Kawasaki, 2009) como *Scppa1*, está anotado en el genoma (versión 9.1) como *Dspp*.

Los elementos reguladores del locus fueron identificados usando datos de ChIP-seq de nuestro laboratorio. Se buscaron las marcas de histonas H3K4me1 y H3K27ac de mayor enriquecimiento en regiones intergénicas e intrónicas cercanas al locus en una ventana de 150pb. También se recurrió a las marcas H3K4me3 y H3K27ac3 que de acuerdo a la literatura son frecuentes en promotores y silenciadores respectivamente.

En principio, marcas de ChIP-seq fueron reconocidos en el locus. Se seleccionaron principalmente aquellos que tuvieran marcas de histonas de mayor enriquecimiento y que correspondían a elementos reguladores putativos. Aquellas regiones que eran ricas en ambas marcas (H3K27ac y H3K4me1) fueron consideradas como MCRs putativos y se evaluaron de forma individual con herramientas bioinformáticas para análisis de secuencia, conservación con

otros genomas, elementos repetidos y otras características que pudieran ser relevantes para el diseño de partidores y el respectivo análisis funcional. Después de dicho análisis se filtraron 4 elementos reguladores que fueron considerados como enhancers putativos. Por su posición en el genoma de *Xt* (Versión 4.1) fueron nombrados como los MCRs 244, 280, 345 y 364, sin embargo, por facilidad de nomenclatura los elementos fueron asignados con las letras del alfabeto A, B, C, D (Ver Figura 24). Cada uno de los elementos seleccionados está río arriba de cada uno de los genes de estudio y son los elementos más proximales a cada gen. El MCR-A es la región de mayor enriquecimiento río arriba del promotor de *Dmp1*, el MCR-B es el único MCR que se encuentra entre *Dmp1* e *Ibsp*, a la vez es el proximal a este segundo gen y por su posición en el genoma corresponde a un elemento intrónico del gen *Mepe*. Finalmente, el MCR-C está ubicado inmediatamente río arriba del gen *Scppa2* a una distancia de 4kk del promotor. Finalmente, debido a su riqueza en ambas marcas de histonas, se consideró el MCR-D, el cuál está en un intrón de otro gen (*Pkd2*) que no forma parte del locus, por lo que se puede considerar un elemento bystander. Debido a que se sabe que los elementos reguladores pueden actuar sobre 1 o varios promotores, no se descartó la posibilidad de que un MRC que está en otro gen pudiera ser un enhancer de alguno de los elementos del locus, por eso fue interesante evaluar también el MCR-D. Otros elementos con enriquecimiento de marca de histonas más alejados, incluso en otra región del locus SCPP (cerca del gen *AMBN* a 200 kb de distancia), no fueron considerados para ensayos funcionales.



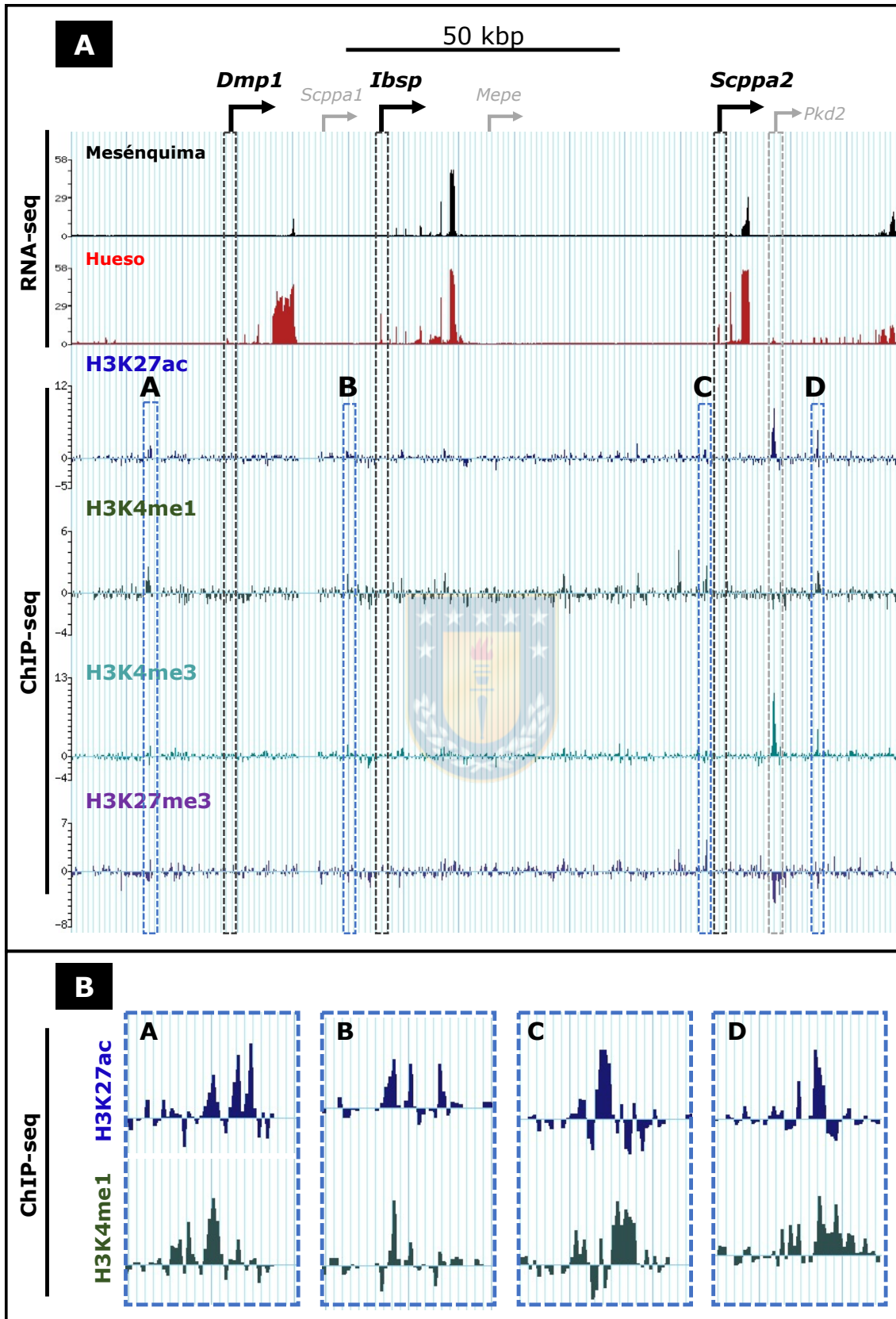


Figura 24: Paisaje regulatorio del locus SCPP

En A se muestra una ventana de 150 kbs que incluye los genes del locus SCPP ácidos. En la región hay 5 genes anotados en la versión 9.1 de *Xt*. Por su orden en el genoma se identificaron como *Dmp1*, *Dspp*, *Ibsp*, *Loc166 (Mepe)* y *Scppa2*. 3 de estos genes son exclusivamente expresados en hueso. El primer exón de cada uno de los genes expresados fue identificado con las cajas negras discontinuas. La región incluye también los enhancers putativos identificados por las marcas de histonas H3K27ac y H3Kme1. La selección de los enhancers putativos incluyó que ambas marcas estuvieran presentes, que las marcas no fueran asociadas a promotores basales (identificados por la marca H3K4me3) de miembros del locus o de otros genes bystander y, en la medida de lo posible, que no presentaran la marca H3K27me3 asociada a enhancers silenciados. Las cajas azules discontinuas señalan en A los enhancers putativos seleccionados (A, B, C, D). Un acercamiento en B de ~2 kbp muestra el enriquecimiento de marcas putativas de enhancer.

Debido a que la modificación H3K27ac correlaciona con MCRs activos mientras que la modificación H3K27me3 correlaciona con promotores silenciados (Bogdanović et al., 2012; Rada-Iglesias et al., 2011), fue importante clonar, en la medida de lo posible, elementos que no tuvieran marcas de histonas de elementos silenciados. Sin embargo, el MCR-C, que es rico en marcas de histonas asociadas a un enhancer putativo fue seleccionado, aunque tenía una región asociada a silenciamiento, comprobado por la marca H3K27me3. A pesar de ello se priorizó su cercanía río arriba del promotor del gen *Scppa2*.

Todos los elementos reguladores seleccionados para ensayos funcionales fueron amplificados y clonados río arriba de cada uno de los promotores basales de los tres genes (ver sección de metodología). Ninguno de los genes del locus SCPP expresados en hueso tiene las marcas canónicas de promotores (H3K4me3 y H3K27ac), al contrario del gen bystander (*Pkd2*), un gen que es particularmente expresado en riñón, pero con niveles muy bajos en hueso. Este último gen tiene las marcas claras de promotor y es visible inmediatamente río abajo del gen *Scppa2* (cajas negras discontinuas en figura B). Por esta observación se concluye que probablemente los promotores de los genes SCPP activos tienen marcas alternativas de histonas.

6.2.3 MCRs putativos ligados al gen *Sparc*

El gen *Sparc* se expresa tanto en mesénquima como en osteoblastos (Figura 23), además de varios tejidos de *Xt* (cerebro, hígado, músculo, riñón, intestino). Su expresión en osteoblastos en pasaje 2, fue también confirmada por RT-PCR (Figura 23). Interesantemente, *Sparc* tiene un paisaje regulatorio enriquecido con marcas de histonas, tanto H3K27ac como H3K4me1 resaltan en una ventana de 120 kb (Figura 25). Es importante observar que tanto río arriba como abajo SPARC está rodeado por dos grandes regiones intergénicas, sin embargo, el enriquecimiento de

marcas asociadas a MCRs putativos se encuentra río arriba. Cinco enhancers putativos fueron identificados río arriba del gen *Sparc* y uno río abajo. Los enhancers fueron determinados y nombrados por su posición en el genoma como MCR-100, 151, 159, 175, 189 y 202. Para una nomenclatura adecuada para este trabajo los elementos fueron designados con una letra del alfabeto (A-F).

Los enhancers putativos fueron analizados por herramientas bioinformáticas. El enhancer B fue particularmente interesante porque coincide con una región altamente conservada además de ser el más proximal al promotor del gen. El MCR F se encuentra río arriba del gen (~60 kb) y es más próximo a otro gen (*Atox1*), un probable gen bystander que no se expresa en hueso. Todos los MCRs fueron amplificados y clonados río arriba del TSS del gen *Sparc* para los ensayos funcionales. Finalmente, al igual que los promotores de los genes del locus SCPP, *Sparc* tampoco tiene las marcas canónicas asociadas al promotor basal.

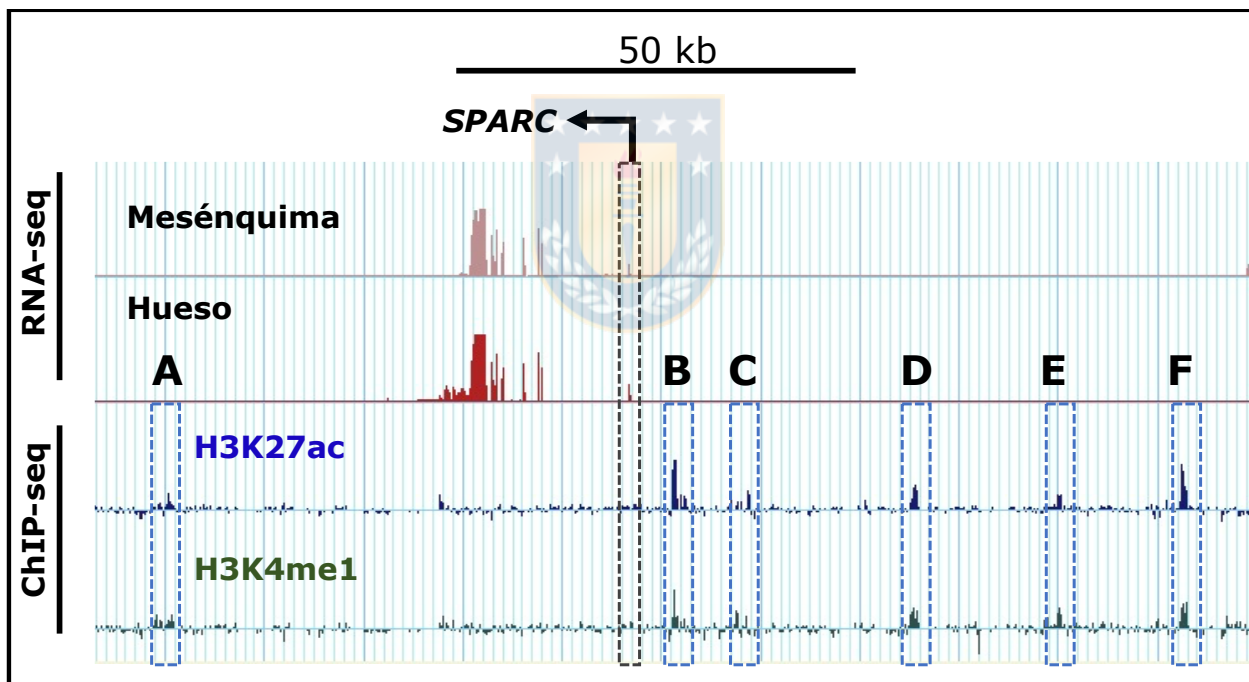


Figura 25: Paisaje regulatorio de *Sparc*

Se muestra una ventana de 120 kbs que incluye dos grandes regiones intergénicas. El primer exón está señalado con la caja negra discontinua. Las cajas discontinuas azules señalan cada uno de los enhancers putativos del paisaje regulatorio identificados por las marcas de histonas H3K27ac (azul) y H3Kme1 (verde).

6.2.4 Un enhancer proximal a *IBSP* activa su expresión en osteoblastos

Para determinar la actividad funcional de los enhancers putativos descubiertos por marcas de cromatina, se procedió a evaluarlos funcionalmente mediante su transfección en cultivo primario de osteoblastos (Figura 26A). Los 4 (A-D) enhancers (~1kb) fueron clonados río arriba del promotor basal del gen *lbsp* (~0.5 kbs). El diseño de plásmidos está esquematizado en la figura 26A. Para todos los ensayos se incluyeron dos controles negativos.

El enhancer proximal al promotor de *lbsp* designado por la letra B (280 por su ubicación en el genoma de *Xt*, versión 4.1) tuvo actividad GFP significativamente mayor que los enhancers C y D. EL enhancer A, que se encuentra río arriba del promotor de *Dmp1*, mostró también diferencia significativa, aunque esta fue menor que el enhancer B. Los enhancer putativos C y D son iguales estadísticamente que los controles negativos TSS y TSS + CNI. La actividad no fue evidente a simple vista en las nubes generadas por el citómetro de flujo, aunque es posible apreciar cómo la población del enhancer B se mueve de una tendencia horizontal a vertical (Figura 26B). Para apreciar mejor la diferencia en las nubes generadas por citometría se extrajeron los datos de fluorescencia de las células individuales y se analizaron estadísticamente (Kruskal-Wallis). En la figura 26B se observa el comportamiento de razón “intensidad GFP / intensidad mCherry” para los diferentes enhancers (en colores) comparados con los controles negativos (cajas grises). Una réplica de este experimento confirmó la actividad del enhancer B, por lo que se concluyó que el MCR280 corresponde a un enhancer proximal del gen *lbsp* mientras que los enhancers A, C y D no tienen actividad sobre el promotor de este gen. Finalmente, debido a que la actividad del enhancer B no es tan evidente como lo será el caso del enhancer 151 sobre el promotor de *Sparc* (ver más adelante), se sugiere que probablemente otros MCRs distales podrían estar activando al promotor de *lbsp*.

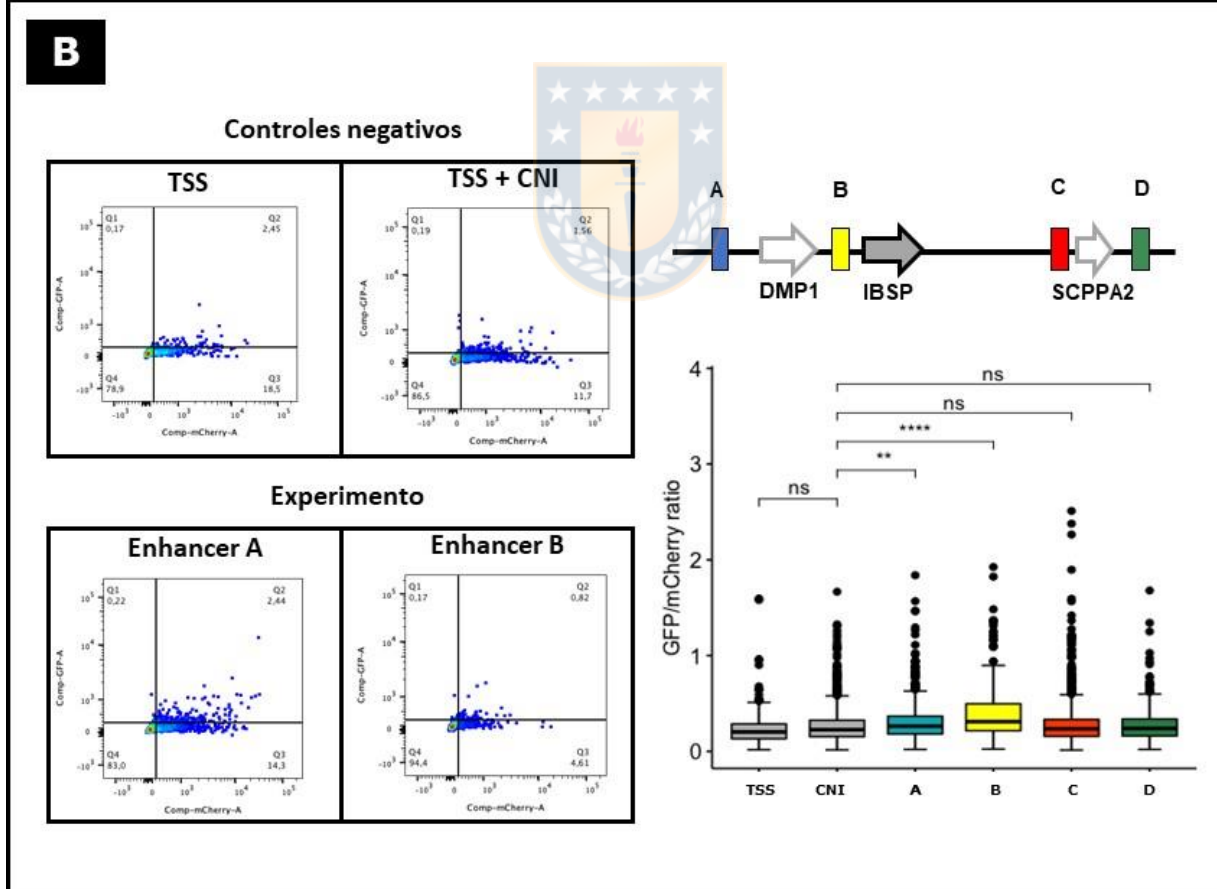
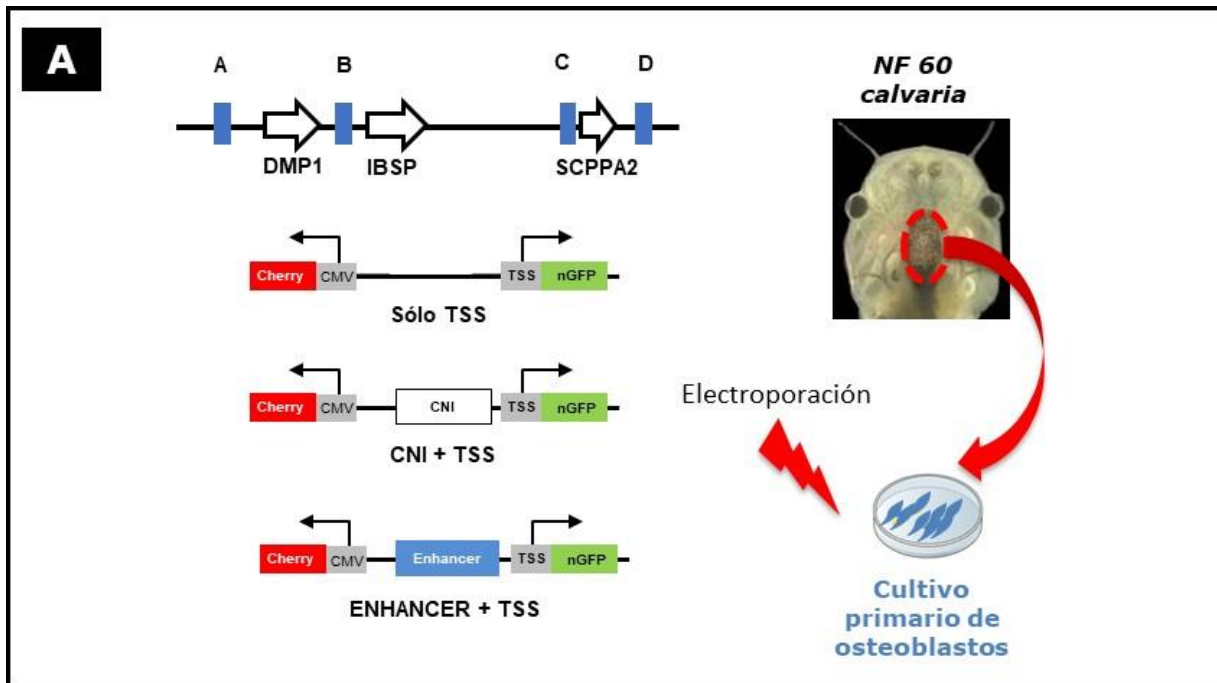


Figura 26: Análisis funcional de MCRs de *Ibsp*

En A se muestra un esquema general del locus SCPP a escala de 150kb. Las cajas azules representan los 4 MCRs probados en este experimento. Abajo se muestran los esquemas de cada uno de los constructos a transfectar. Cada MCR fue clonado en plásmidos independientes. La actividad transcripcional de cada MCR fue cuantificada en función de la intensidad de la proteína verde fluorescente (GFP). Las células de cultivo primario de osteoblastos fueron extraídas de del hueso frontoparietal (calvaria) de renacuajos de *Xt* (NF58-60). Las células fueron expandidas hasta un pasaje 2 y fueron transfectadas por electroporación. La fluorescencia fue cuantificada por citometría de flujo. En B se muestran los gráficos generados para cuatro constructos (TSS-CNI-MCR-A y MCR-B). En los gráficos cada cuadrante representa poblaciones celulares con diferentes valores de fluorescencia (roja en el eje X y verde en Y). Los controles suelen tener una mayor cantidad de valores de fluorescencia en el cuadrante rojo, dado que son células positivas transfectadas. Si un MCR activa más que el control, suelen tener mayores valores de fluorescencia en el cuadrante Q2, debido a que representan células electroporadas positivas rojas (vertical) que también tienen fluorescencia verde (horizontal). No siempre es fácil evaluar a simple vista si un MCR es activo, por ello se recurre a extraer cada valor de fluorescencia. Los datos se normalizan mediante la razón GFPn/mCherry, lo que permite realizar estadística de comparación de medias, así como visualizar la distribución de los datos en gráficos de cajas. En este experimento los MCRs A y B muestran densidades de poblaciones con valores de media mayores respecto a los controles, lo que se asocia a una mayor activación del promotor. El MCR-B es el de mayor significancia, lo que lo determina como un enhancer proximal de *Ibsp*. Todos los experimentos fueron realizados con al menos dos réplicas experimentales y 2 réplicas biológicas. Los datos de fluorescencia fueron extraídos con el software R y se aplicó la razón GFPn/mCherry para normalizar. Para todos los análisis estadísticos se aplicó el método no paramétrico Kruskal-Wallis y una corrección con el método estadístico de Bonferroni. Valores de significancia ns = $p > 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

6.2.5 Dos MCRs proximales al promotor de *Scppa2* muestran actividad silenciadora

Se seleccionaron 3 enhancers putativos (A, C y D) que fueron clonados río arriba del promotor basar del TSS del gen *Scppa2* dirigiendo la expresión de la proteína fluorescente GFP (ver sección de metodología) para evaluar su actividad funcional. Sorprendentemente, los enhancers putativos A y D tuvieron una actividad silenciadora del promotor de *Scppa2*, algo que no fue visto para *Ibsp* (sección anterior). Dicho comportamiento era probablemente esperado para el MCR-C, dado que es el único que presentó la marca represiva de histonas (H3K27me3), este último fue estadísticamente similar que los controles negativos. Desafortunadamente, el enhancer B, que activó el promotor de *Ibsp* no logró ser clonado corriente arriba del TSS de *Scppa2*. No obstante, fue interesante evaluar los dos enhancers proximales a *Scppa2*, lo cuáles no activaron al promotor en un contexto de osteoblastos.

Lo controvertido de este experimento nos llevó a replicarlo al menos 3 veces por el sistema de transfección por electroporación Neon en diferentes cultivos obteniendo siempre el mismo resultado. Aunque el proceso de estandarización de la electroporación nos llevó a establecer condiciones que no afectaban la viabilidad celular, se pensó que el alto voltaje podría

afectar la identidad celular y por ello afectar el resultado, tratándose el silenciamiento de los enhancers A y D artefactos derivados de la técnica. Por lo anterior, se decidió optar por otro método de transfección (sistema *FuGene*) que, aunque es menos eficiente que la electroporación, podría tener la ventaja de eliminar el posible efecto negativo de la electroporación sobre la identidad celular. Sin embargo, el resultado fue exactamente el mismo.

Con este resultado se concluye que los MCRs A y D tienen una actividad represiva de la transcripción del gen *Scppa2* en cultivo primario de osteoblastos, mientras que el MCR-C, proximal a su promotor, no afecta la regulación transcripcional.

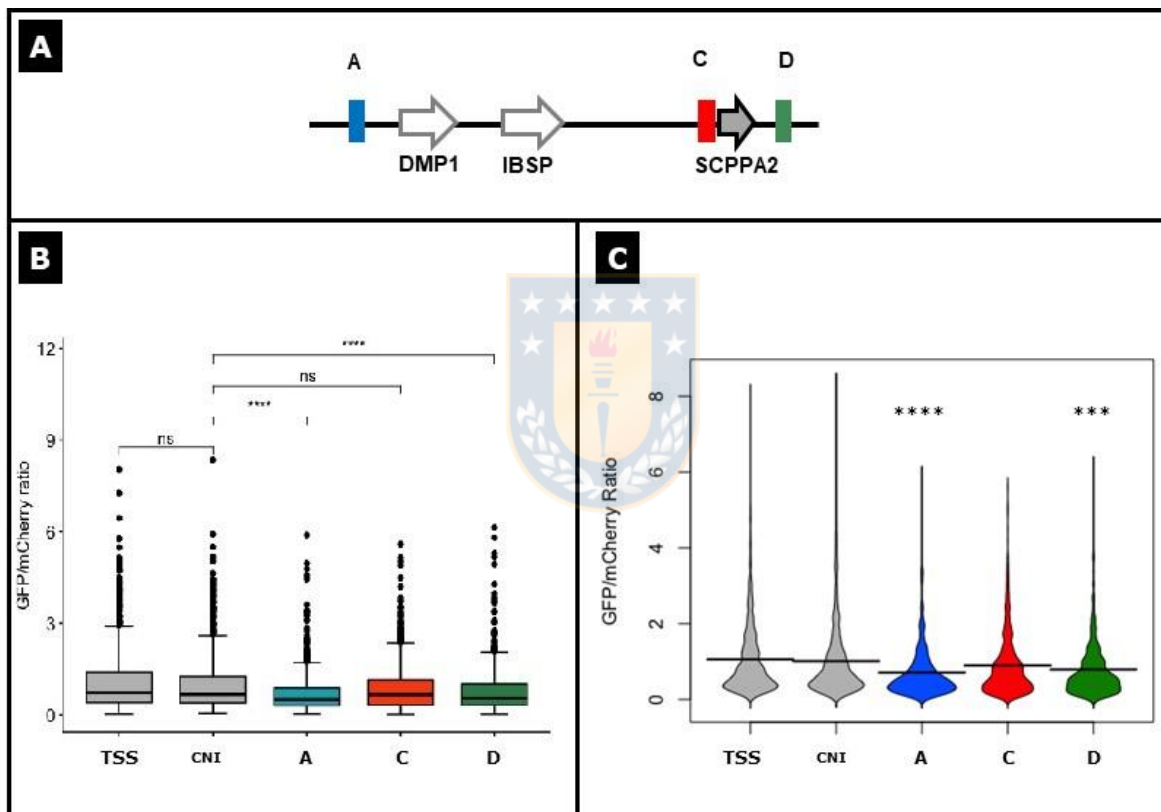


Figura 27: Análisis funcional de MCRs de *SCPPA2*

En A se muestra un esquema a escala de 150kbs del locus SCPP, cada caja de colores corresponde a cada uno de los MCRs probados en este experimento. Los MCRs C y B se encuentran río arriba y abajo respectivamente. El MCR-C es proximal (4kb) al promotor de *Scppa2* (gen representado con flecha gris). En B se muestra la comparación de poblaciones de datos de fluorescencia (GFPn/mCherry) generados por citometría de flujo. Los controles negativos comparados entre ellos no muestran diferencia significativa, sin embargo, los MCRs A y D son significativos respecto a los controles, y muestran una razón de fluorescencia menor, lo que puede sugerir una actividad silenciadora del promotor. El MCR-C proximal y enriquecido fuertemente por marcas de histonas no fue diferente de los controles. Los resultados para este

promotor fueron replicados 3 veces con diferentes cultivos celulares en pasaje 2 cuantificados 48 h después de la electroporación. Este experimento también fue hecho mediante transfección por sistema *Fugene* y los datos tuvieron el mismo comportamiento. El análisis estadístico se realizó mediante el test de Kruskal-Wallis y una corrección con el método estadístico de Bonferroni. Valores de significancia ns = $p > 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

6.2.6 Los enhancers putativos del locus SCPP no alteran la expresión de *Dmp1*

Ningún enhancer seleccionado del locus SCPP afecta la expresión del gen *Dmp1*. Un resultado que era esperado dado que por RT-PCR e hibridación in situ se demostró que *Dmp1* es un gen que se activa exclusivamente en osteoblastos. Era muy importante cerciorarse de la actividad de los MCRs putativos en este gen porque *Dmp1* se activa rápidamente en células que diferencian a osteocitos. Este experimento sirvió como control negativo para validar que los MCRs estudiados en ensayos funcionales fueron evaluados exclusivamente en un contexto de osteoblastos y que la actividad del enhancer B de *Ibsp* y B de *Sparc* (ver la siguiente sección), así como de los silenciadores de *Scppa2* fueron propios del contexto transcripcional de osteoblastos y no sólo elementos artefactuales, lo que le da validez a nuestros resultados. Sería interesante replicar el experimento en células que se induzcan a diferenciación de osteocitos y/o que inicien procesos de mineralización, algo que no fue parte de los objetivos de este trabajo pero que puede sugerirse como proyección.

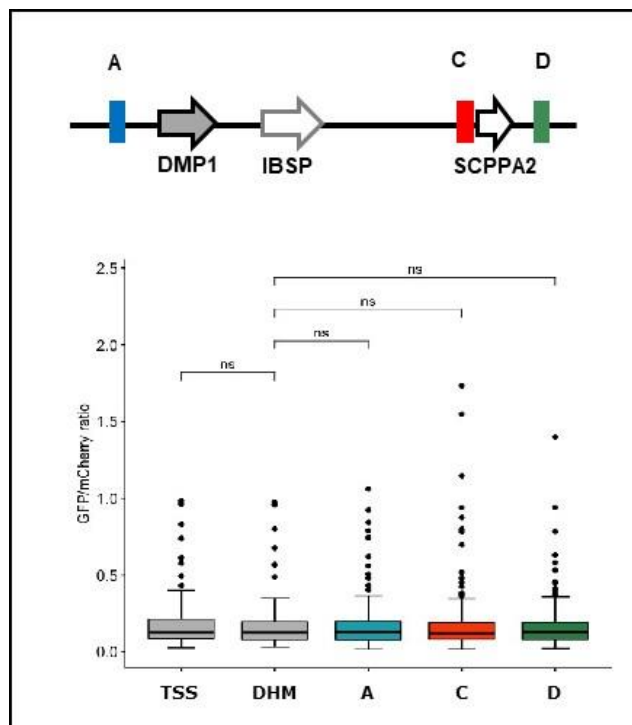


Figura 28: Análisis funcional de MCRs de *Dmp1*

Se muestra un esquema a escala de 150 kbs del locus SCPP, cada caja de colores corresponde a cada uno de los MCRs probados en este experimento. Los MCRs C y B se encuentran distales al promotor de *Dmp1*. El MCR-A se encuentra a 17kb del promotor (gen representado con flecha gris). La comparación de poblaciones de datos de fluorescencia (GFPn/mCherry) generados por citometría de flujo muestra que no hay diferencia significativa entre los MCRs y los controles de *Dmp1*. Los resultados para este promotor fueron replicados 3 veces con diferentes cultivos celulares y en diferentes pasajes (1, 2 y 3) cuantificados 48 h después de la electroporación. El análisis estadístico se realizó mediante el test de Kruskal-Wallis y una corrección con el método estadístico de Bonferroni. Valores de significancia ns = $p > 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

6.2.7 Dos MCRs proximales al promotor de *Sparc* activan la transcripción

Los enhancers putativos mostrados en la sección 2.3 fueron amplificados, clonados en diferentes plásmidos ENTRY que contenían el promotor basal de *Sparc* y transfectados por electroporación en cultivo primario de osteoblastos (Figuras 29 A B y C). Tres enhancers distales (>40 kb) fueron similares a los controles negativos, lo que implica que no tuvieron una actividad significativa sobre el promotor de *Sparc*. Los enhancers A, B y C presentaron actividad en este ensayo. Interesantemente el enhancer A se encuentra a una distancia de 50 kb y presenta una región conservada con el genoma de *X. laevis*. Sin embargo la actividad del enhancer A no es tan potente como el caso del enhancer B, éste último presentando uno de los mejores ejemplos de activación de promotores de todos los experimentos realizados con ensayos reporteros. Su activación pudo ser perceptible incluso en los gráficos de células cuantificadas por citometría de flujo. Se observa una clara distribución en diagonal, debido a que las células fueron positivas tanto en el cuadrante de mCherry como en el de GFP (Figura 29C). Es muy importante mencionar que una parte de la región clonada está evolutivamente conservada. Otro aspecto relevante a considerar es que la región del enhancer B contacta físicamente a su promotor, tal como será detallado en la siguiente sección. El enhancer C también tuvo una interesante activación de su promotor en células de pasaje 2, aunque el experimento estadísticamente es significativo, sólo una población celular fue activada (Figura 29D), dicha observación puede aseverarse por la media poblacional (línea negra en el gráfico de cajas).

Todos los experimentos con reporteros presentados en la figura 29 fueron llevados a cabo en el pasaje 2 y evaluados 48 hrs después de su transfección. Cuando se intentó replicar el experimento con los enhancers B y C, el enhancer C dejó de mostrar actividad mientras que el enhancer B continuó siendo activo. Este hallazgo llevó a confirmar la actividad del enhancer B pero surgió la duda respecto al enhancer C, cuya actividad pudo haberse tratado simplemente

de un elemento artefactual. Esto llevó a preguntarse si el pasaje celular podría estar afectando la actividad de dichos enhancers, por lo que se intentó realizar el experimento de los enhancers B y C en el pasaje 1, pensando en que esto sería cercano a las condiciones de osteoblastos nativos. Debido a que se había estandarizado la electroporación en 1×10^5 células, fue posible replicar el experimento en el pasaje 1. Sorpresivamente, el enhancer C, que fue positivo una de tres veces en el pasaje 2 (ver tabla 3) fue claramente activo en el pasaje 1 mientras que el enhancer B conservó su actividad en los pasajes 1 y 2 (ver tabla 3). Tomando en cuenta esta dinámica se decidió electroporar los enhancers B y C en el pasaje 3. El enhancer B que había sido activo en los pasajes 1 y 2 se inactivó en el pasaje 3. Desafortunadamente no se lograron recuperar suficientes células para evaluar el enhancer C en pasaje 3.

De todo lo anterior se concluye la actividad funcional del enhancer B, un elemento proximal (~4 Kb) al promotor de SPARC. Además, se sugiere que la actividad de los enhancers es probablemente dependiente del pasaje celular. Estas observaciones sugeridas no están respaldadas por una N contundente (excepto para el enhancer B), estudiar la actividad de enhancers en diferentes pasajes celulares no era un objetivo de este proyecto, por lo tanto, se puede sugerir como una proyección.

En conclusión, al menos tres MCRs putativos tuvieron actividad de enhancer en el paisaje regulatorio de SPARC. Los enhancers B y C fueron los de mayor significancia estadística respecto a los controles negativos y se replicaron al menos 3 veces. El enhancer A sólo fue confirmado 1 vez y sólo en el pasaje 2.

	ENHANCER A	ENHANCER B	ENHANCER C
PASAJE 1	No probado	2 / 2	2 / 2
PASAJE 2	1 / 1	2 / 2	1 / 2
PASAJE 3	No probado	0 / 2	No probado

Table 3: Resumen de la actividad funcional de los enhancers A,B y C de SPARC en diferentes pasajes

Se muestran los resultados de todos los experimentos realizados con los reporteros de SPARC con la fórmula X / Y. X representa el número de veces que el experimento fue positivo y Y el número de veces que fue probado experimentalmente. El enhancer B fue positivo en todos los experimentos realizados en los pasajes 1 y 2, sin embargo, el enhancer C sólo fue positivo 1 vez en el pasaje 2 y dos veces en el pasaje 1. En el pasaje 3 sólo el enhancer B fue probado pero su actividad fu similar a los controles negativos.

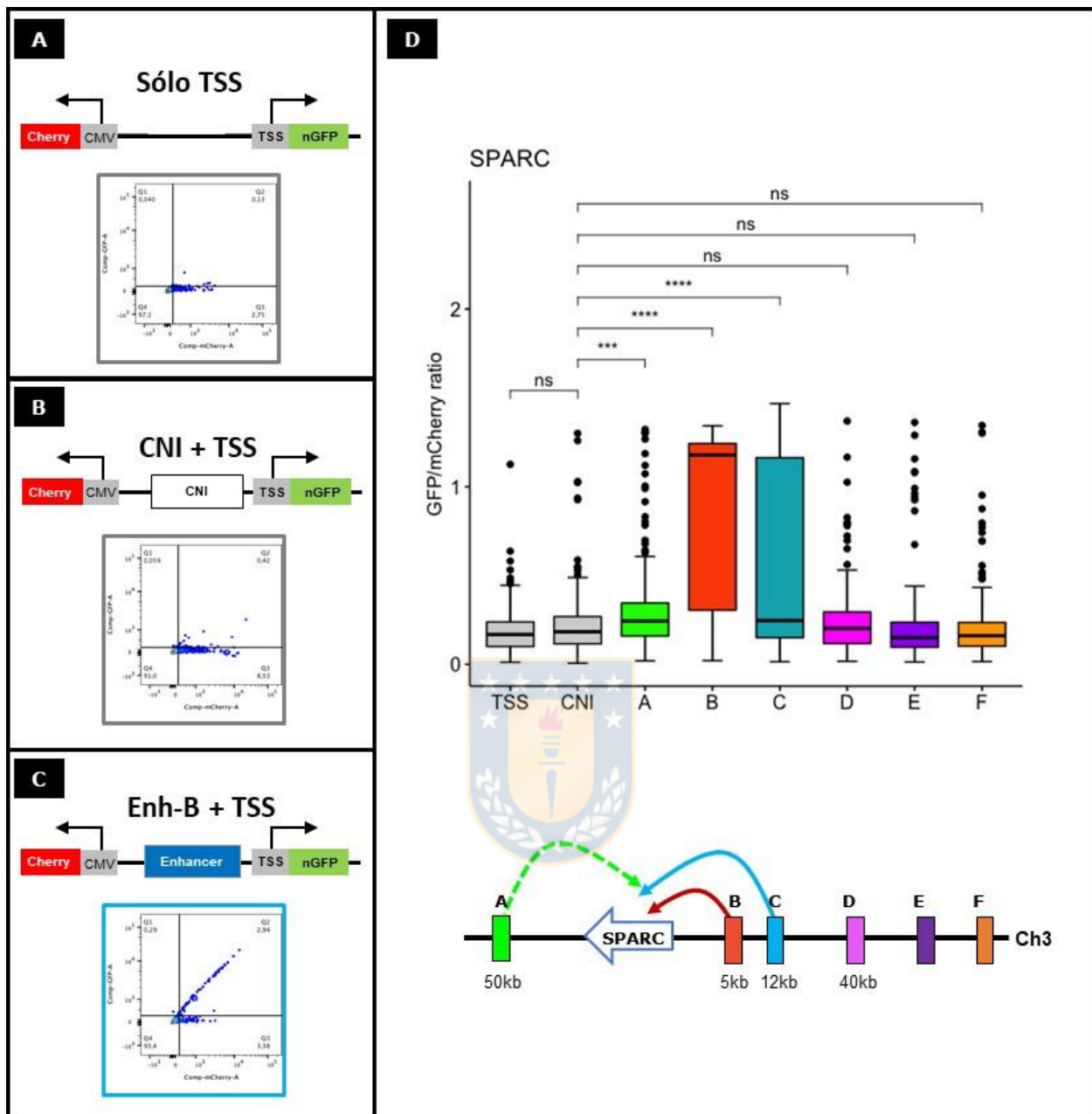


Figura 29: **Análisis funcional de MCRs de Sparc**

En A, B y C se muestran los esquemas de constructos para cada condición: dos controles y el ejemplo de un MCR que activó la expresión de GFP. Los gráficos derivados de citometría muestran el comportamiento de datos en cada uno de los cuadrantes. En C, el resultado de un MCR activo muestra una diagonal que implica que una población celular tiene fluorescencia roja y verde (Q2). Los datos extraídos y graficados mediante la razón (GFP/mCherry) muestran en gráficos de cajas la actividad positiva de al menos 3 potenciales enhancers (A, B y C). El enhancer B es el más proximal al promotor de SPARC y es el que fue más significativo. Otros MCRs distales (>40kb) fueron similares a los controles. Un esquema explicativo de una región de 120 kb que incluye las dos grandes regiones intergénicas de SPARC muestra la distribución de los diferentes MCRs. Las flechas muestran los MCRs que activan el promotor de SPARC,

B y C con mayor fuerza que el MCR-A. Los controles junto con los MCRs B, C y F se replicaron 3 veces en diferentes pasajes (1 y 2) cuantificados 48 h después de la electroporación. EL MCR-A, D y E sólo tienen una réplica en la cuál se analizaron todos los MCRs en un solo experimento. El análisis estadístico se realizó mediante el test de Kruskal-Wallis y una corrección con el método estadístico de Bonferroni. Valores de significancia ns = $p > 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

6.3 Perfil de interacciones de los miembros SCPP y *Sparc*

Los genes SCPP son esenciales en los procesos de mineralización de hueso, dentina y esmalte de vertebrados. Tres genes SCPP se expresan fuertemente en hueso de *Xt*. En secciones anteriores se demostró que *Dmp1* tiene un patrón de expresión diferencial entre osteoblastos y osteocitos a pesar de su cercanía física a *Ibsp* y *Scppa2* en la secuencia de ADN. Probablemente esto se debe a elementos reguladores que activan su transcripción en un contexto osteocítico.

A lo largo del locus SCPP hay enhancers putativos que fueron identificados por las marcas de la histona 3. Sin embargo, tal como se observó en la sección anterior, no es fácil determinar si las marcas de histona proximales a los genes son los elementos clave que dirigen la actividad transcripcional en osteoblastos. Se sabe que pueden existir elementos reguladores que participan en conjunto, por lo que la actividad de un solo enhancer no es suficiente para activar la transcripción. Además, aunque un porcentaje de genes tienen elementos *cis*-reguladores proximales existen otros distales que han sido menos estudiados funcionalmente y que también pueden controlar la transcripción vía contactos físicos de la cromatina en el espacio entre los enhancers y los promotores. En otras palabras, regiones alejadas en el genoma pueden interactuar con uno o más promotores blanco para regular su transcripción.

La hipótesis de este trabajo fue que posibles enhancers pueden estar contactando diferentes miembros del locus SCPP corregulando su expresión en un contexto osteogénico. Para poner a prueba nuestra hipótesis, el objetivo 3 consistió en aplicar la técnica de Captura de Conformación Circular 4C con la finalidad de estudiar regiones puntuales de cromatina que sirvieron como carnadas (*baits*) para atrapar regiones del genoma que están próximas, distales o incluso en otros cromosomas. Las regiones utilizadas como carnadas fueron los promotores basales de cada uno de los miembros del locus SCPP expresados en hueso. De esta forma, se pudo determinar el paisaje regulatorio de cada uno de los miembros del locus. En esta sección se detallarán los resultados de las secciones de interacción en *Cis*. Las interacciones en *Trans* (con otros cromosomas) se encuentran en el anexo 2 aunque no se discute al respecto.

6.3.1 Paisaje de contactos del gen *Ibsp*

El pipeline del paquete *r3Cseq* generó gráficos de interacción entre regiones del genoma y cada uno de los promotores estudiados. Después de un proceso de normalización, fue posible asignar valores estadísticos a las regiones de interacción significativa. Así se obtuvieron todos los posibles fragmentos que interactuaron con su promotor en osteoblastos de cultivo primario.

En la figura 30 se muestran las regiones de interacción cercanas del *Vp* (+500 y -500 kb) de *IBSP*. En la parte superior de la imagen se muestra una barra gris que corresponde a una gran sección del cromosoma 1 de *Xt*. Cada una de las líneas verticales corresponde a una región de interacción, coloreada en función de los valores *q-value*. Una mayor intensidad de color rojo representa una mayor probabilidad de interacción.

La región de 1 MB alrededor del *Vp* de *Ibsp* se señala con un triángulo rojo invertido y se amplifica (líneas negras discontinuas) para mostrar los contactos más significativos marcados por círculos rojos oscuros (*q-value* \geq 0.00001). Para reconocer cada una de las principales regiones que contactan al promotor de *Ibsp* se asignaron números romanos (I-VII) a las regiones con contactos de mayor *q-value*. En el eje de las “y” se señalan la cantidad de *reads* normalizados que enriquecen cada uno de los *peaks* mostrados.

El análisis en *r3Cseq* generó también gráficos denominados *dominiogramas* para visualizar el enriquecimiento local de contactos entre fragmentos de cromatina. La función utiliza el valor score de la probabilidad de contacto en ventanas de 2 a 25 kb. El gráfico utiliza colores que van del gris al rojo para representar el nivel de interacción. Cada línea vertical en el eje de las X representa una sección de 1 kb del cromosoma y cada rectángulo horizontal, indica el tamaño de la ventana. Por ejemplo, las regiones II y III que se encuentran a ~230 y ~130 Kb respectivamente de distancia río abajo del promotor presentan altas frecuencias de interacción, así como regiones dentro de los primeros 100 kb río arriba del promotor. Los dominiogramas son útiles para comparar cómo cambia la frecuencia de interacción entre dos condiciones, una experimental y otra control. Aunque no se realizaron los experimentos de 4C en otros tejidos “controles”, si se evaluaron varios promotores de genes próximos en el cromosoma, por lo que la función *domainogram* sirvió para comparar contactos significativos entre los miembros del locus.

En la figura 30 se anotaron los genes asociados a la región genómica mostrada. Se representan los genes SCPP ácidos y los ricos en Pro/Gln (polígonos azules y rojos respectivamente) así como los diferentes miembros vecinos o genes bystander (polígonos grises). La mayor parte de regiones que contactan al promotor de *IBSP* se encuentran en los

primeros 100 kb río arriba del promotor ($1e+05$), abarcando los genes *Dmp1*, *Odam* y *Cdk12* mientras que dos regiones de contactos distales (B y C) se encuentran río abajo, en regiones intergénicas e intrónicas de otros genes.

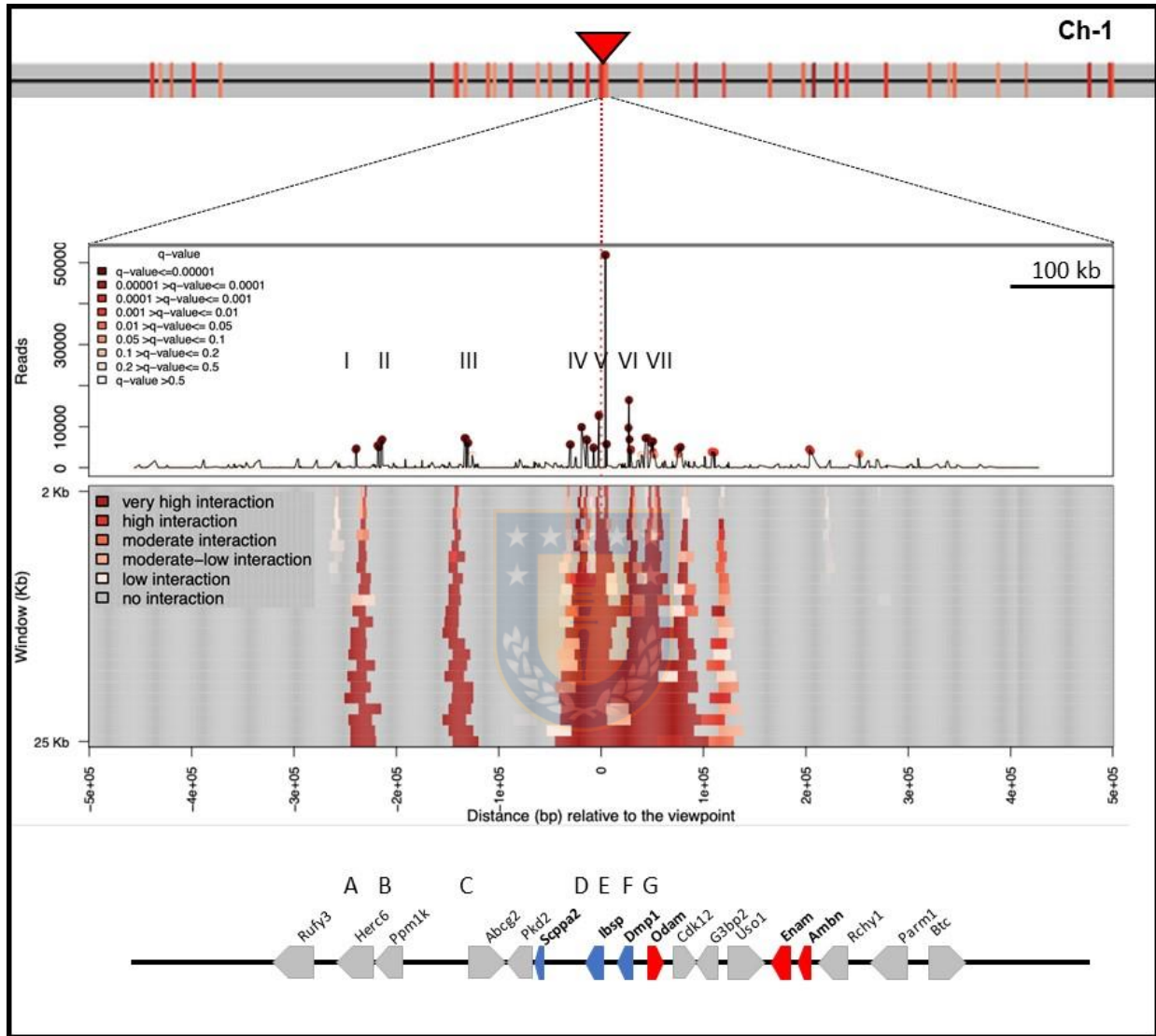


Figura 30: Perfil de interacciones de *lbsp*

Se muestran las regiones de interacción candidatas con el promotor de *lbsp* en osteoblastos exploradas en 3 tipos de gráficos. El gráfico generado con la función *plotOverviewInteractions* del paquete *r3Cseq* exploró la distribución de regiones de interacciones a escala genómica. En este perfil se muestra sólo una región del cromosoma 1 donde se encuentra el locus *SCPP* (otros cromosomas pueden observarse en el anexo 2). Las regiones rojas oscuras representan las regiones de mayor interacción con el promotor. El *viewpoint* se señaló en el cromosoma con un triángulo invertido rojo. Con la función *plotInteractionsNearViewpoint* se realizó un gráfico de aumento (+500/-500 Kbs) alrededor del *Vp* para visualizar la distribución de regiones de interacción en una ventana de 1MB. Esta función genera las

referencias de los genes, los fragmentos de restricción (no mostrados) así como las señales candidatas de mayor interacción señaladas con un círculo que va de colores pálidos hasta el rojo oscuro. Las referencias de genes son generadas por el software y se esquematizan abajo para una mejor visualización (azul = SSCP ácidos, rojos = ricos en Prolina (Pro) y Glutamina (Gln) y, los genes grises = elementos bystander). Las regiones de mayor frecuencia de interacción fueron señaladas con los números romanos (I-VII). Otra función generó los gráficos *dominiogramas*, los cuáles muestran las señales de interacción detectadas utilizando un análisis basado en *ventanas* que van de 2 a 30 kb. Los gradientes de color de los dominiogramas representan la intensidad de la señal de interacción detectada para cada ejecución (run) de la ventana definida (valor de q transformado). Los dominiogramas fueron generados usando los parámetros por default asignados por el software. Los gradientes de color representan la cobertura (*coverage*) normalizada de la ventana.

Se realizó un acercamiento en la región del *Vp* de *Ibsp* (Figura 31). Se muestra una ventana de 12 kb en donde se observan dos regiones de interacción significativa con el promotor. La primera región (derecha) coincide con un intrón del gen anotado en el *genome browser* de la UCSC como *Dspp* (*Scppa1*). Interesantemente, esta región está próxima al enhancer B, el cuál fue significativo en ensayos reporteros. La segunda región (izquierda) es la de mayor significancia de interacción con el promotor de *Ibsp* y abarca el último exón de *Scppa1*. Además, usando datos preliminares de ATAC-seq en osteoblastos de nuestro laboratorio, encontramos que hay regiones de cromatina abierta que coinciden con ese contacto. Esa región también se encuentra conservada en la evolución de vertebrados.

De todo lo anterior se concluye que el extremo 3' del gen *Scppa1* interactúa fuertemente con el promotor del gen *Ibsp*. La actividad funcional de una región cercana a este contacto fue comprobada mediante ensayos reporteros.

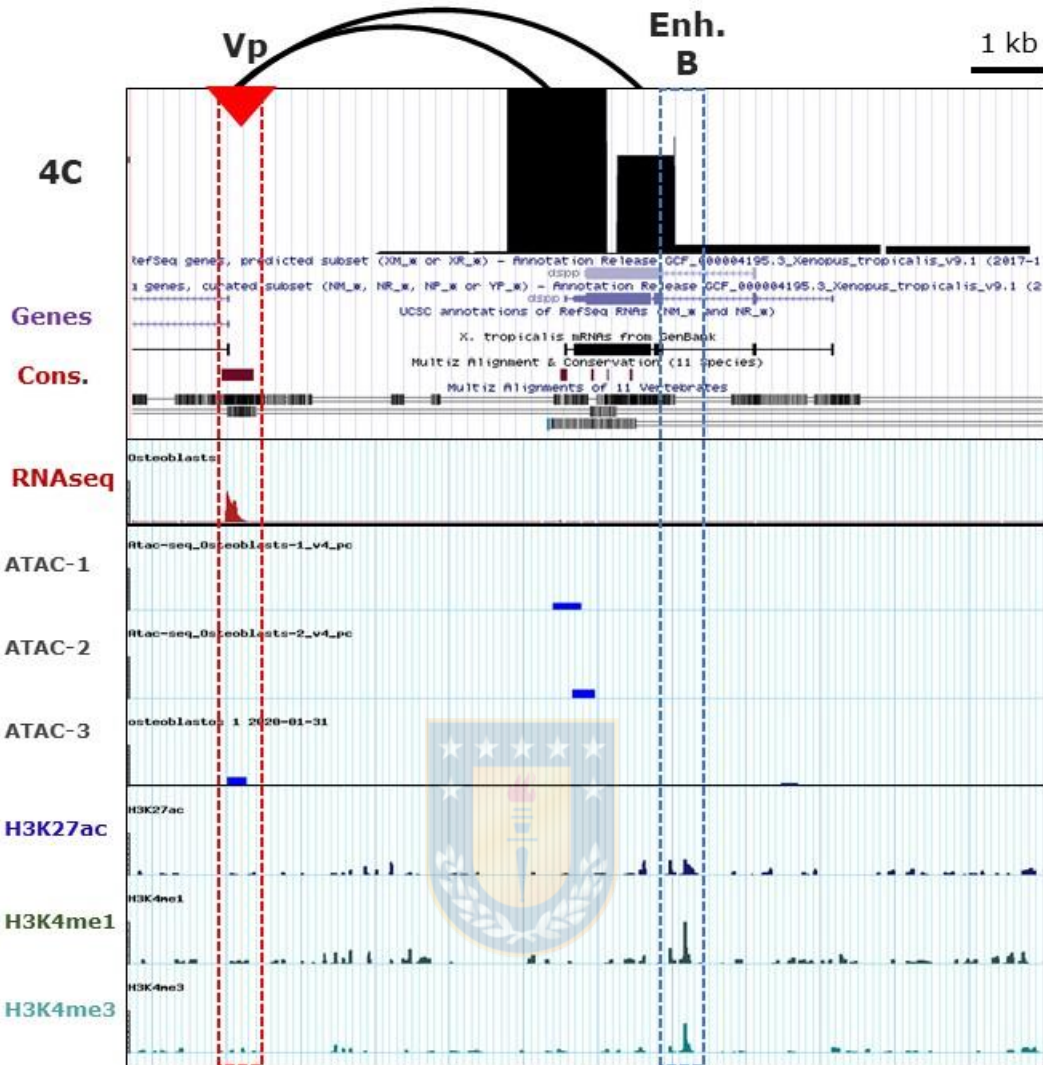


Figura 31: Interacción proximal al promotor de *IBSP*

La interacción de mayor significancia estadística se encuentra a 5 kb de distancia del promotor ($q\text{-value} > 0.0001$). Para evitar ruido, las señales de interacción muy potentes generadas siempre en 4C adyacentes al *Vp* fueron excluidas. Los archivos *bigwig* fueron cargados al *genome browser* de la UCSC para su visualización en el genoma de *X. tropicalis Jul. 2016 (Xenopus_tropicalis_v9.1/xenTro9)*. La región de mayor interacción se encuentra en el extremo 3' del gen *Scppa1*. Esta región mostró también una alta señal de cromatina abierta en dos réplicas de ATAC-seq realizadas en nuestro laboratorio. Un contacto contiguo coincide también con una región intrónica rica en marcas de histonas y que fue probada en los ensayos reporteros de la sección anterior, la cuál mostró actividad sobre el promotor de *Ibsp* (líneas discontinuas azules). El promotor de *Ibsp* (líneas discontinuas rojas) muestra señal de cromatina significativamente abierta en una réplica de ATAC-seq que coincide exactamente con el primer exón (gen anotado, conservación y primer exón determinado por RNA-seq).

Todas las regiones que interactuaron de forma significativa con el promotor de *Ibsp* se resumen en la figura 32. Si se cruzan los datos ChIP-seq con 4C, pueden destacarse algunos aspectos relevantes. El contacto I se encuentra en una región intrónica del gen *Herc6*, el intrón más cercano tiene marca de H3K4me1. Respecto al contacto II, interesantemente coincide con un intrón del gen *Ppm1k* que es rico en H3K4me1 y H3K4me3. Los contactos de la región III caen en un intrón del gen *abcg2*, son conservados pero se encuentran completamente libres de marcas de histonas. Los contactos en la región IV se encuentran en el extremo 3' del gen *Ibsp*, pero carecen de conservación y de marcas de histonas. El contacto V fue explicado en la sección anterior (mostrado en la imagen 31). El contacto VI es muy interesante porque justo se encuentran en la región promotora de *Dmp1*, donde hay completa ausencia de marcas de histonas, lo que es acorde con la inactividad de este promotor en osteoblastos. En la región VII, un contacto significativo coincide con el MCR-A (244), el cuál tuvo actividad parcial sobre el promotor de *Ibsp* en ensayos reporteros. Es importante mencionar que el contacto VII se encuentra justamente en la región promotora del gen *Odam*, el cuál pertenece al locus SCPP-P/Q que no se expresa en hueso. Probablemente, la región VII se encuentre en otro dominio regulatorio separado por un insulador, lo que imposibilitó que dicho MCR tuviera fuerte actividad sobre el promotor de *Ibsp*. Además, esta región tuvo actividad silenciadora sobre el promotor de *Scpp2* (Figura 27).

6.3.2 *Ibsp* contacta al promotor del *Dmp1*

En la sección anterior se mostró el enriquecimiento de interacciones significativas con el promotor de *Ibsp*. Aunque no se detallaron las frecuencias de interacción del promotor de *Dmp1*, es importante mencionar que una proporción de regiones que contactan al promotor de *Ibsp* son compartidas con *Dmp1*. Estas interacciones incluyen la del promotor de *Ibsp* con la región promotora de *Dmp1* y viceversa (Figura 33).

Cuando se realizan experimentos de 3C (uno contra uno) y un promotor contacta con un enhancer, se realiza el mismo experimento inverso, es decir se usa como *Vp* el enhancer, así se comprueba que efectivamente ambas regiones interactúan y no se trata sólo de un elemento artefactual. En nuestros resultados, fue sumamente relevante realizar el 4C utilizando como *Vp* el promotor de *Dmp1* ya que esto nos permitió tener una idea del enriquecimiento de contactos de este gen en osteoblastos, así como corroborar la interacción entre los promotores *Ibsp-Dmp1*.

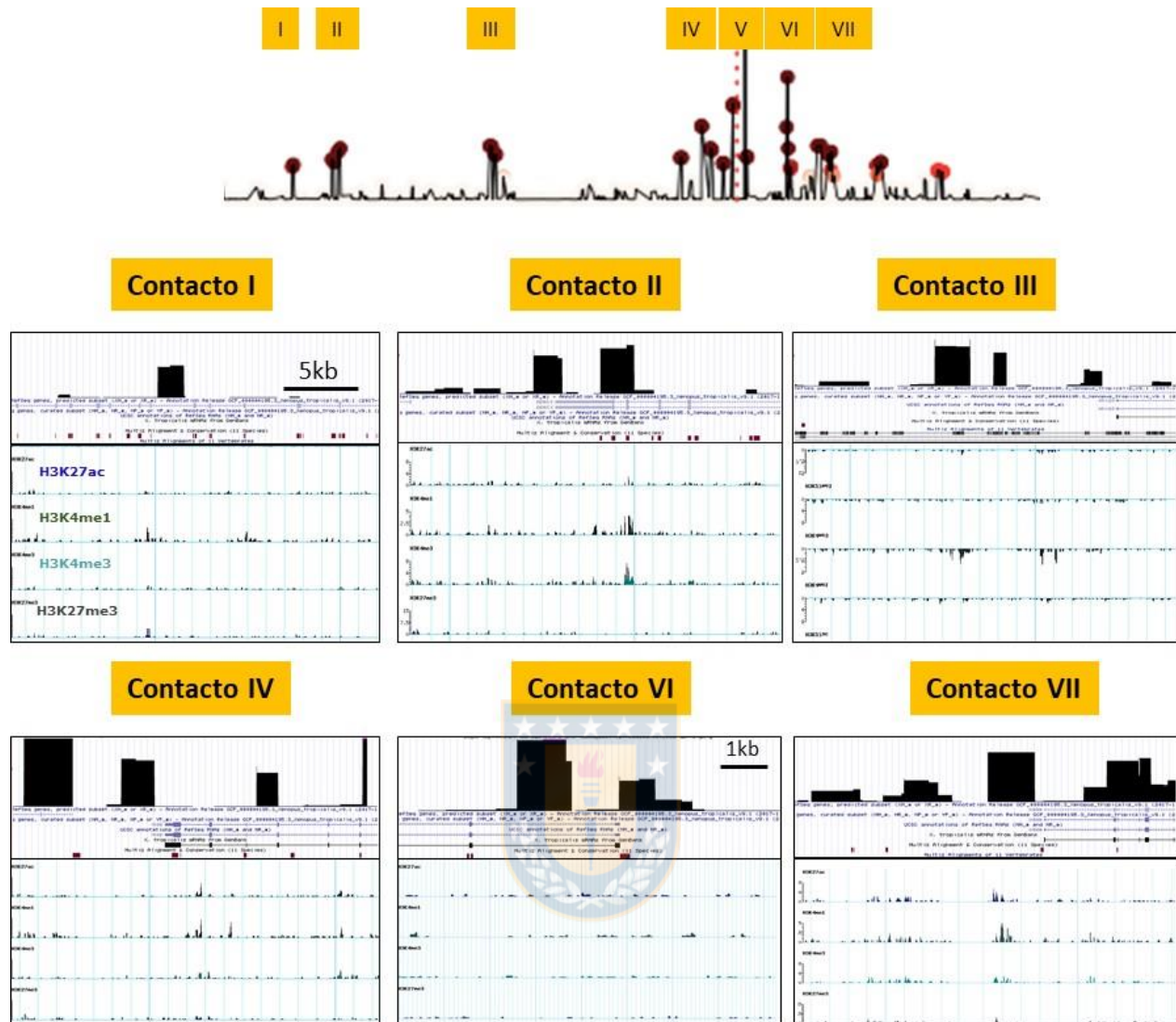


Figura 32: Otras regiones candidatas de interacción con el promotor *lbsp*

Las regiones de interacción significativa del ensayo 4C de *lbsp* fueron analizadas en el *genome browser* y comparadas con regiones de marcas de histonas. Los contactos I, II y VII caen en regiones enriquecidas en marcas de H3 ($q\text{-value} > 0.0001$). Interesantemente, el contacto VI' y VI'' coinciden con el primer intrón y el promotor respectivamente del gen *Dmp1* y están completamente libres de marcas de histonas, lo que concuerda con un gen que no se expresa en osteoblastos. El contacto principal en VII, coincide precisamente con la región promotora de *Odam*, un gen SCPP que está apagado en hueso. Interesantemente, esta región contiene al MCR-A que por un lado activó parcialmente a *lbsp* y por el otro, silenció al promotor de *Scppa2* en ensayos reporteros.

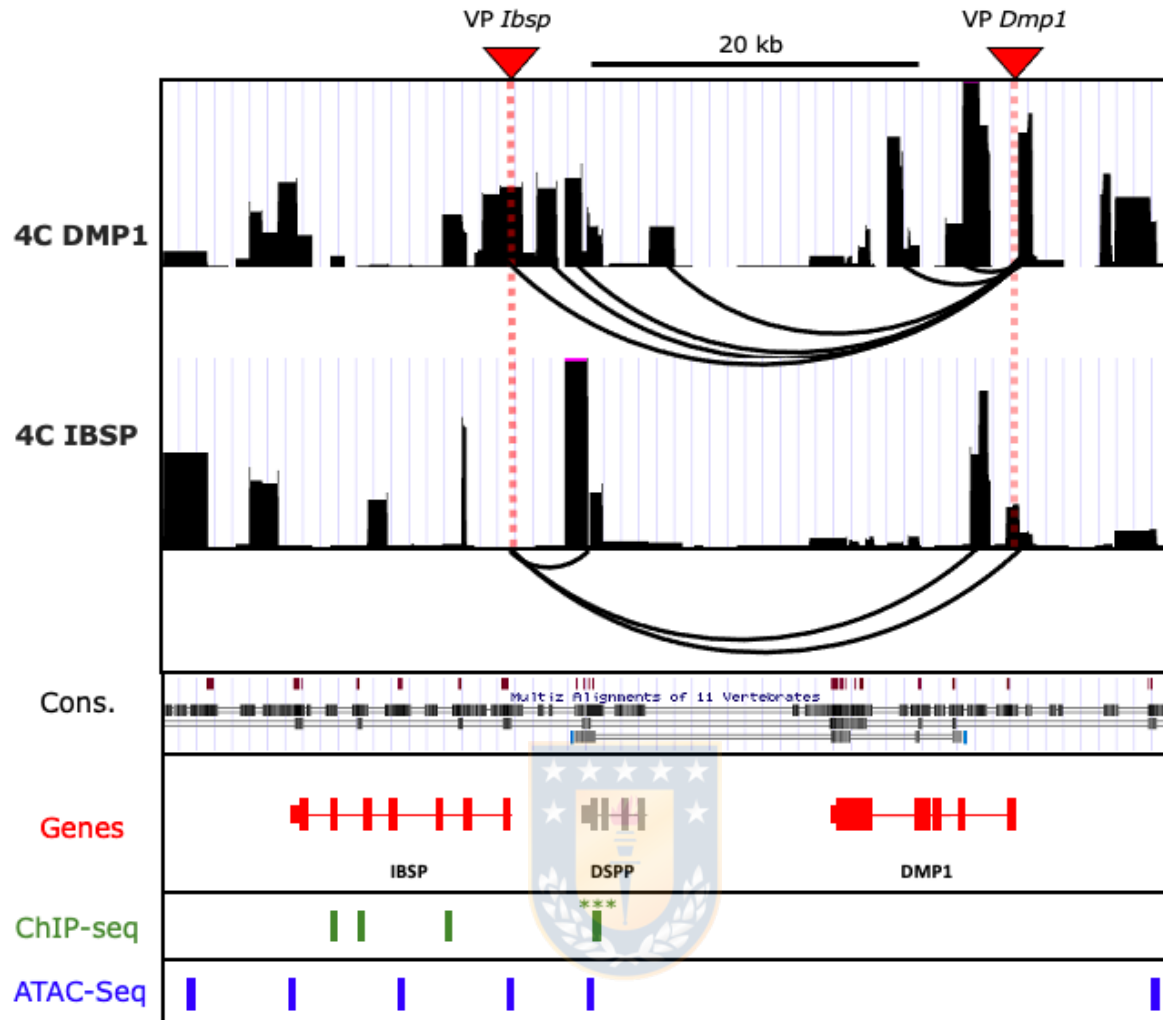


Figura 33: Los perfiles de interacción sugieren que los promotores de *Ibsp* y *Dmp1* se contactan

Las regiones promotoras de *IBSP* y *DMP1* muestran regiones ricas de interacción entre si. La línea roja punteada (arriba) muestra los VP de ambos genes. Las marcas enriquecidas de marcas de histonas están representadas con barras verdes mientras que la cromatina abierta con barras azules (abajo). El elemento proximal de contacto a *IBSP* muestra las marcas de histonas (los asteriscos muestran que este enhancer fue funcional en ensayos reporteros) y coincide con la región de interacción y cromatina abierta. Es importante observar que en la región del cuerpo génico de *Dmp1* no hay enriquecimiento de marcas de histonas o señales de cromatina abierta pero aún así se sugieren los contactos fuertemente con la región promotora de *IBSP* (q-value > 0.0001).

6.3.3 Paisaje de interacciones del gen *Scppa2*

En este proyecto se estudiaron también las frecuencias de interacción del promotor de *Scppa2*. Este gen se expresa en osteoblastos y osteocitos y está inmediatamente río abajo de *Ibsp*. En la figura 35 se observa que la mayor parte de interacciones se encuentran alrededor de 100kb río arriba y abajo del promotor con la excepción de un fuerte contacto a 150 kbs río arriba (derecha), la cuál presenta amplias regiones de interacción (hasta 25 kb) que se encuentran entre los genes bystander *Cdk12*, y *G3bp2*.

Otro aspecto relevante es que este promotor interactúa débilmente en *cis* con el cromosoma 1, a diferencia de lo que se observa con *Ibsp* (ver barra de interacción cromosomal en la parte superior de las figuras 30 y 35). También se puede observar menor probabilidad de interacción con otros cromosomas (anexo 2).

El promotor de *Scppa2* interactúa fuertemente con tres regiones ubicadas en los intrones de *Pkd2*, un gen bystander que conserva sintenia con el locus *SCPP* a lo largo de la evolución de vertebrados. *Pkd2* es un gen que se expresa principalmente en riñón. En el cromosoma 4 del genoma humano, *Pkd2* se encuentra en sintenia con el gen *Scp1*, éste último comparte la misma ubicación cromosomal que *Scppa2*.

En vertebrados *Pkd2*, *Scppa2* y *Spp1* (gen relacionado a *Scppa2* en cuanto a estructura y función) se expresan en riñón mientras que *Scppa2* y *Spp1* son exclusivos de hueso. Tomando en cuenta la relación posicional y funcional entre *Spp1* de humano y *Scppa2* de *Xenopus*, y dadas las interacciones de éste último con *Pkd2*, es posible entonces sugerir porqué ambos elementos no se separaron en la evolución durante millones de años y siguen conservando sintenia. Posiblemente hay elementos regulatorios en los intrones del bystander *Pkd2* que son funcionales sobre el promotor de *Scppa2*. Un posible modelo que explica este hecho será sugerido en la discusión.

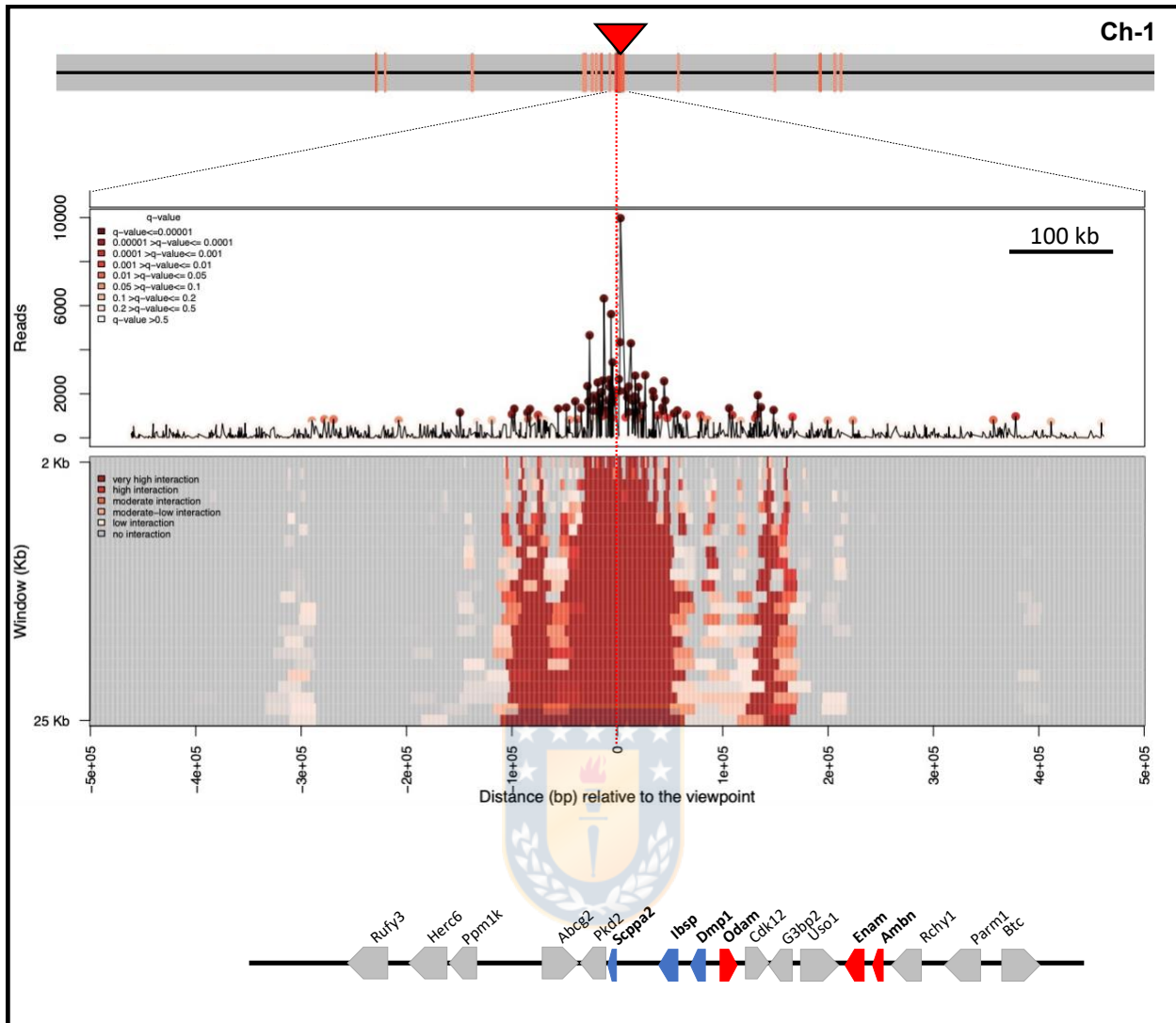


Figura 34: **Perfiles de interacción de *Scppa2***

Se muestra la distribución de las regiones de interacción candidatas con el promotor de *Scppa2*. Una región del cromosoma 1 muestra que este promotor interactúa débilmente en *Cis*. El *Vp* se señaló en el cromosoma con un triángulo invertido rojo. Un aumento de la región del *Vp* permite visualizar la distribución de regiones de interacción en una ventana de 1 MB. Las referencias de genes son esquematizadas abajo para una mejor visualización (azul = SCPP ácidos, rojos = ricos en Prolina (Pro) y Glutamina (Gln) y, los genes grises = elementos bystander). Las regiones de mayor frecuencia de interacción pueden observarse alrededor de 100 kb del *Vp*, el *dominiograma* confirmó este comportamiento de interacción proximal al *Vp*. Los gradientes de color de los *dominiogramas* representan la intensidad de la señal de interacción en ventanas de 2 a 25 kb. Una región distal de alta frecuencia de interacción fue detectada a 150 kb río arriba del promotor (q-value > 0.0001).

Si se observan de cerca algunos contactos de *Scppa2* es interesante destacar que a diferencia de *Dmp1*, *Scppa2* no comparte regiones de contacto con el promotor de *Ibsp* (Figura 37). No obstante, hay un fuerte contacto con la región 3' de este gen (líneas negras discontinuas). Dicho contacto está conservado y muestra señales de cromatina abierta (ATAC) así como marcas de histonas en la vecindad del contacto. Es importante mencionar que el promotor de *Scpp2* no interactúa con el extremo 3' de *Scppa1*, a diferencia de *Ibsp* y *Dmp1*, quizá esta separación, pueda explicar un subdominio de este gen, lo que probablemente también sea un argumento a favor del patrón de expresión diferencial de *Scppa2* (Figura 36). Sin embargo, el fuerte contacto con el intrón de *Ibsp* y con los genes baystander río arriba del promotor puedan explicar porqué en la evolución este gen sigue compartiendo sintenia y no se movió a otras regiones del genoma.

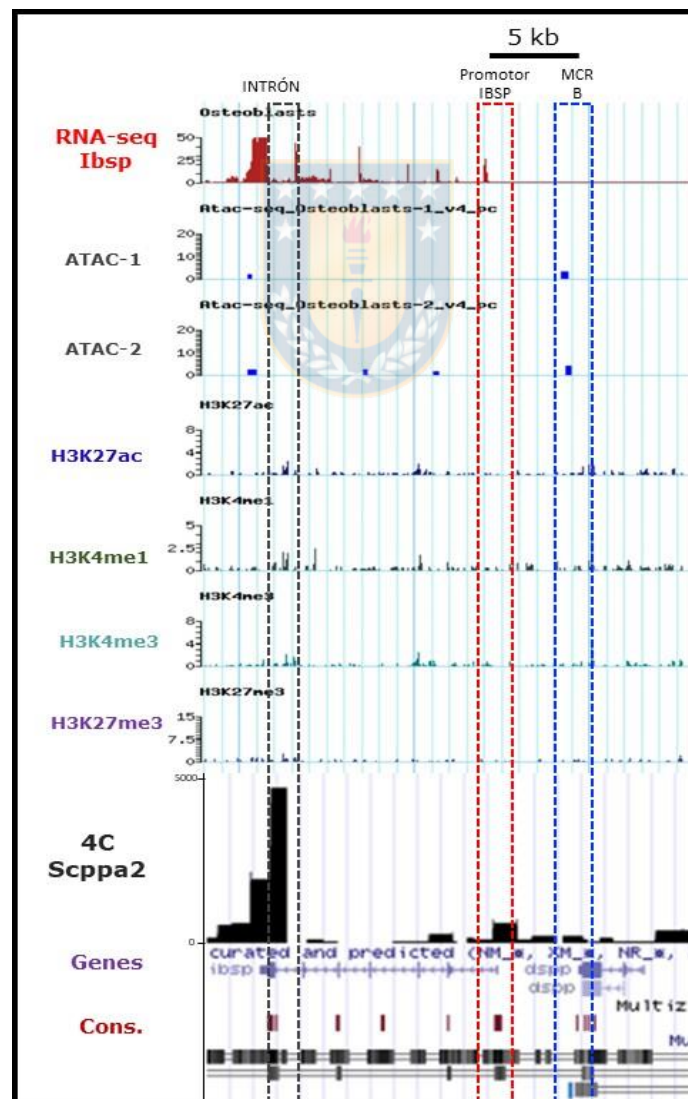


Figura 35: El promotor de *Scppa2* interactúa con el último exón de *Ibsp*

Uno de los contactos significativos de *Scppa2* río arriba del promotor (q-value > 0.0001) coincide con el extremo 3' de *Ibsp*, interesadamente en el último intrón, el cual tiene una región conservada, así como marcas de histonas. Esta región de interacción es específica y no es enriquecida ni en el promotor de *Ibsp* ni en el último exón de *Scppa1*, a diferencia de lo que sucedió con los promotores de *Ibsp* y *Dmp1*.

Los MCR-C (345) y MCR-D (364) utilizados para ensayos reporteros no activaron la transcripción del promotor de *Scppa2* en el contexto de osteoblastos, por el contrario, el MCR-D tuvo actividad silenciadora (Figura 27). Esto se debe probablemente a posibles regiones de unión para factores de transcripción que pueden reprimir la transcripción. Cuando se analizaron estas regiones cruzando con los datos de 4C fue posible observar que el MCR-D se encuentra en uno de los intrones de *Pkd2* que tuvieron alta señal de interacción, aunque no hay señales de regiones abiertas de cromatina para estos intrones. Respecto al MCR-C se observa que no tiene alta frecuencia de interacción con el promotor de *Scppa2*, además de que no cuenta con señales de cromatina abierta (ATAC-seq). Sin embargo, hay una señal de interacción significativa justo al lado del MCR-C. Llama la atención que esta región sea una de los que tienen mayor número de *reads* y que su tamaño abarque casi 5kb, además de tener dos grandes regiones conservadas con *Xenopus leavis*. Sería interesante evaluar funcionalmente dicha región, así como el primer intrón de *Pkd2* cuyo valor de interacción (q-value) es también significativo (Figura 37). Se diseñaron partidores para clonar y evaluar estas regiones (ver archivo LADE 2020-11-10-Partidores-4C) y se sugiere como una actividad de proyección.

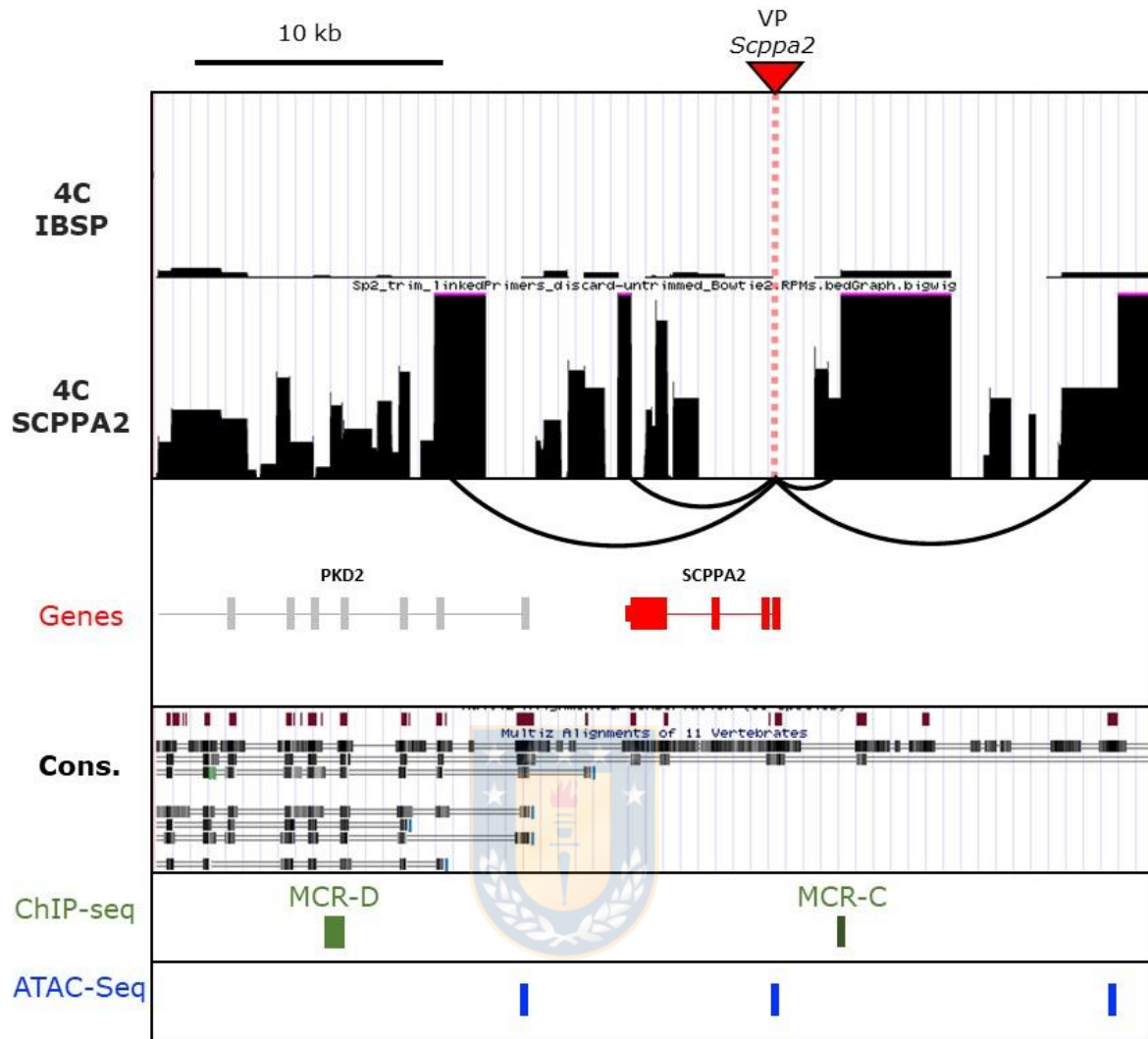


Figura 36: ***Scppa2*** contacta significativamente los intrones del gen bystander ***Pkd2***

Valores significativos de interacción proximales a *Scppa2* (q -value > 0.0001) coinciden con las regiones intrónicas del gen *Pkd2*, un gen que se expresa en riñón, pero no en hueso. En uno de estos intrones (4^o) se encuentra el MCR-D probado en experimentos reporteros y que tuvo actividad silenciadora cuando fue clonado río arriba del promotor de *Scppa2*. El MCR-C es proximal al promotor de *Scppa2*, pero no coincide con el contacto de mayor interacción, aunque está cercano. Interessantemente la región de mayor interacción coincide con dos regiones conservadas con ranas. La señal de cromatina abierta coincide con los promotores de ambos genes, así como con otro contacto significativo de *Scppa2* que además también se encuentra conservado con *X. leavis*.

6.3.4 Paisaje de interacciones del gen *Sparc*

Uno de los objetivos de esta tesis fue evaluar el paisaje regulatorio del gen *Sparc*. Primero, por su origen evolutivo que es compartido con el gen *Sparc1*. De este último, se sabe que surgieron, por duplicaciones en tándem, los SCPP. Segundo, porque contiene regiones enriquecidas con marcas de cromatina. El análisis funcional por ensayos reporteros de estas regiones mostró que los MCRs A, B y C presentaron actividad en un contexto de osteoblastos. Y tercero, porque *Sparc* es un gen semi-ubícuo que se expresa en una gran cantidad de tejidos, particularmente en hueso, donde se mostró que se expresa en osteoblastos con remanentes mensajeros en osteocitos, de forma muy similar al gen *Ibsp*.

El análisis de interacciones mostró que *Sparc* es uno de los promotores con altas frecuencias de contactos en *Cis* con el cromosoma 3 (Figura 38A). Sin embargo, si se estudia la región de 1MB alrededor del VP, se observan las mayores regiones de enriquecimiento en una zona muy cercana al promotor. De hecho, varias regiones de alta frecuencia se encuentran repartidos en el primer intrón de *Sparc*, así como alrededor de los primeros 10 Kb río arriba del promotor y sólo algunas regiones significativas distales río arriba del promotor (~200kb).

En la figura 39 se muestran las interacciones principales de *Sparc* en una región de 100 Kb alrededor del promotor. Es posible ver en el recuadro de líneas rojas discontinuas al menos 7 regiones significativas. Si se miran detalladamente las regiones proximales en el recuadro ampliado (abajo), se pueden observar el primer intrón rico en contactos así como una región de aproximadamente 10 Kb río arriba del promotor parcialmente enriquecida de regiones de interacción. Interesantemente el primer intrón de *Sparc* tiene tres regiones conservadas, una de ellas con ratón y las otras con especies de rana justamente donde las interacciones tienen mayor valor de significancia estadística. Una señal fuerte sobresale a ~4kb río arriba del promotor, esta región está justo al lado de la única región conservada proximal y es muy cercana al enhancer B (líneas discontinuas azules), el elemento que fue probado de forma exitosa por ensayos reporteros. Otro MCR que fue funcional es el enhancer C, el cuál se encontró entre dos regiones de contactos parciales pero aún significativas.

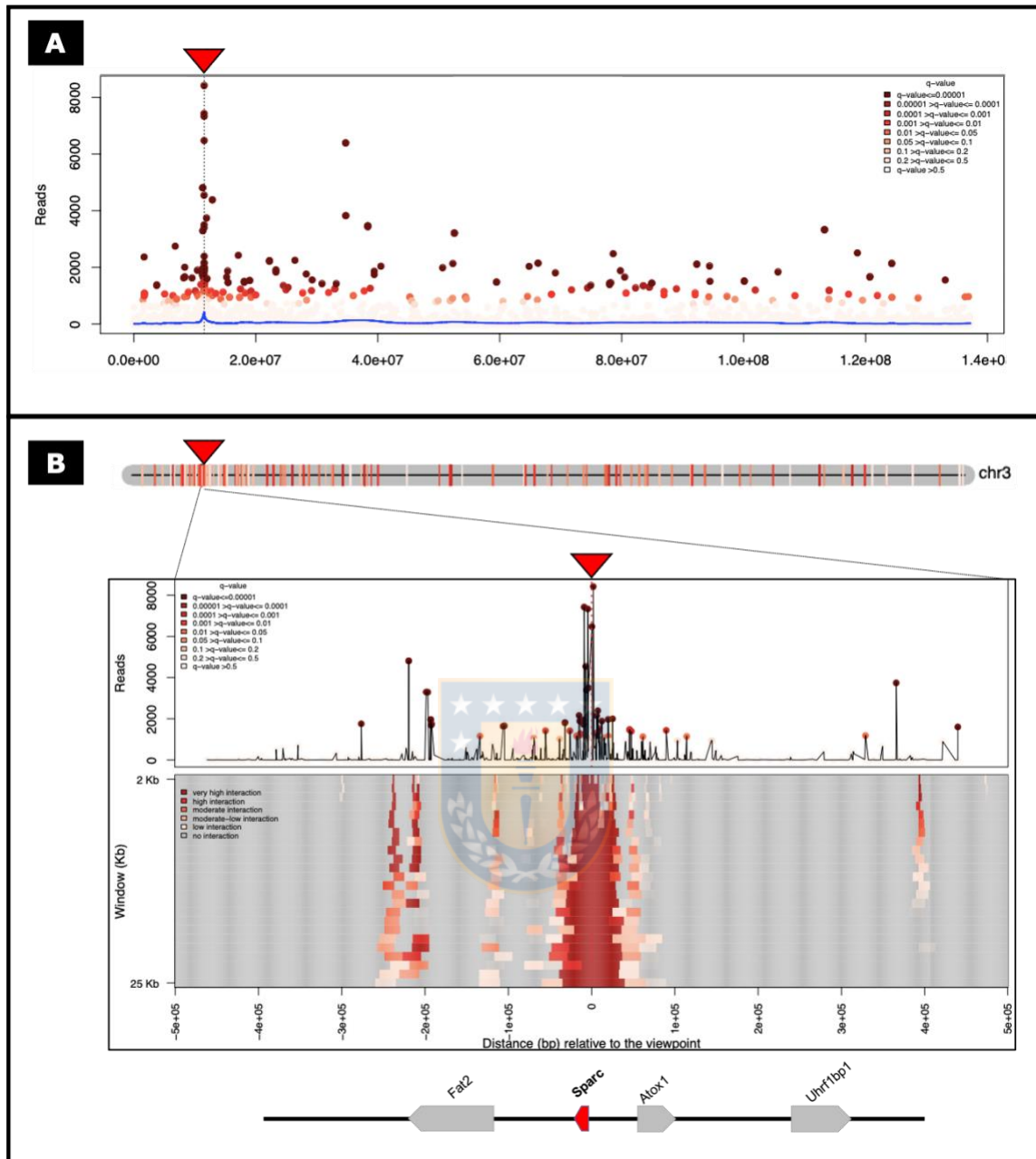


Figura 37: Perfiles de interacción de SPARC

En A se generó un gráfico con la función *plotInteractionsPer-Chromosome* que muestra las regiones de interacción a lo largo del cromosoma en *cis*. El *viewpoint* se señaló en el cromosoma con un triángulo invertido rojo. La mayor parte de interacciones significativas se encuentran alrededor del VP, sin embargo, un par de interacciones a distancia pueden percibirse a ~23Mb. En B (arriba) se muestran los mismos datos a lo largo del cromosoma representando cada interacción significativa con colores que van del gris (nula interacción) al rojo (mayor interacción). Un acercamiento de la región del VP permite visualizar la distribución de regiones de interacción en una ventana de 1MB. El gen SPARC y los genes más próximos se esquematizan abajo. Las regiones de mayor frecuencia de interacción pueden observarse alrededor de

20 kb del *Vp*. No obstante algunos contactos representativos alcanzan la periferia de los genes bystander *Atox1* y *Fat2*. Los gradientes de color de los dominiogramas representan la intensidad de la señal de interacción en ventanas de 2 a 25 kb. Al menos dos regiones distales de alta frecuencia de interacción fueron detectadas a ~200 kb río arriba del promotor (q-value > 0.0001).

Alrededor de los contactos descritos anteriormente no hay regiones de contactos significativas, tampoco hay regiones conservadas, esta observación abarca los MCRs putativos que fueron probados funcionalmente (C, D, E, F) y que no activaron la transcripción.

Con lo anterior se concluye que probablemente en osteoblastos de *X. tropicalis*, la arquitectura de la cromatina en el locus de *Sparc* genera una región rica de contactos proximales que permiten la llegada de factores de transcripción que trabajan conjuntamente para activar la expresión de este gen en osteoblastos. Además, los datos de ATAC-seq permiten confirmar el hecho de que tanto el promotor como la región proximal al promotor estén abiertas. Quizá esta región de ~20kbs genera un microdominio de asociaciones que a su vez permiten que otras regiones del cromosoma 3 interactúen fuertemente con esta sección en la proximidad espacial, lo que generó que varias secciones del cromosoma 3 quedarán capturadas tal como se muestra en la figura 38A y en la figura 5 del anexo 2.

Finalmente, es importante destacar que las interacciones distales de *Sparc* están en la periferia de los genes bystander *Atox1* y *Fat2*. Esto es muy relevante porque ambos genes en humano se encuentran en los límites de un TAD, cada uno en un micro-dominio. El punto esquinado que separa estos genes deja a *Fat2* y *Sparc* en un subTAD, lo que podría explicar los contactos distales entre *Sparc* y el 3' de *Fat2* (Figura 37 y Anexo 4). Este hallazgo da validez a nuestros resultados y podría explicar la conservación de sintenia de estos genes durante millones de años de evolución de vertebrados. Probar la funcionalidad del contacto distal de *Fat2* es una de las proyecciones de este proyecto. Se diseñaron los partidores para clonar y evaluar estas regiones (ver archivo LADE 2020-11-10-Partidores-4C).

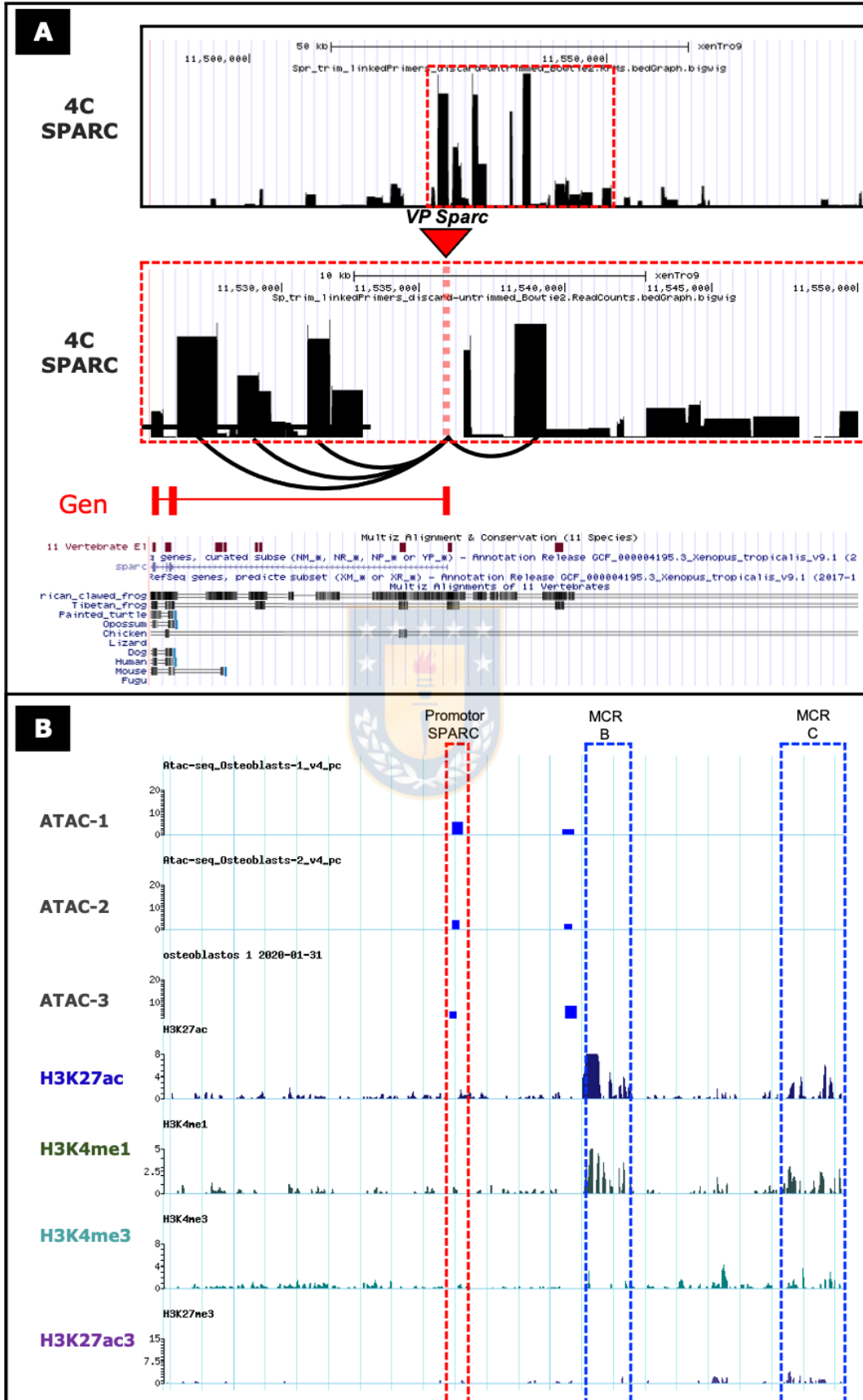


Figura 38: **Perfiles de interacción proximales al promotor del gen *Sparc***

En A se muestran las regiones de contacto proximales al promotor de SPARC (q-value > 0.0001) en una ventana de 100 Kb, la región de líneas rojas discontinuas es ampliada y el VP es señalado con un triángulo rojo invertido. Las líneas negras unen las regiones que interactúan con el promotor y permiten señalar al menos tres regiones significativas de interacción en el primer intrón de *Sparc* y en una región proximal río arriba del promotor (~4 Kb). En B, se cruzan los datos de ChIP-seq y ATAC-seq con 4C. El promotor de *Sparc* muestra marcas de cromatina abierta pero no de histonas. A 4 Kb del promotor se muestra otra región de cromatina abierta que coincide con una región conservada en el genoma. Muy cerca de esta región se encuentra la señal de interacción principal (a la derecha) así como fuertes señales de H3K27ac y H2K4me1 (a la izquierda) que corresponden con el MCR-B que fue activo en ensayos reporteros. Río arriba, justo a 12 Kb del promotor se observa el MCR-C, justo en la región donde termina la zona principal de contactos del promotor de *Sparc*.

6.3.5 Todos contra todos

Para evaluar las regiones ricas de interacciones compartidas entre todos los promotores se alinearon todos los dominiogramas (Figura 40). El 4C también fue realizado con el Vp del promotor de la *Ameloblastina* (*Ambn*) cuyo gen no se expresa en hueso, pero forma parte del locus SCPP P/Q. Interesantemente, las regiones ricas de contactos de este promotor se ubicaron relativamente proximales al Vp con escasos contactos compartidos con los SCPP ácidos. Esta separación en dos claros dominios podía deberse a la distancia del locus SCPP ácido respecto al promotor *Ambn* o por la arquitectura de la cromatina, es decir, una mayor frecuencia de interacciones entre los elementos SCPP ácidos derivados de su agrupación en un dominio, GRB o TAD.

A grandes rasgos se puede observar cómo los genes *Dmp1*, *Ibsp* y *Scppa2* comparten regiones de interacción, sin embargo, otras, como la principal región de contactos de *Ibsp* que son compartidos con *Dmp1*, divergen parcialmente de *Scppa2*.

Algunas regiones de interacción son también específicas de cada promotor, como es el caso de la que está río arriba del promotor de *Dmp1* (derecha), así como dos claras regiones río abajo de *Ibsp* (izquierda), una compartida con *Scppa2* y otra única de *Ibsp*.

Con los dominiogramas alineados y los genes anotados (abajo) pueden determinarse algunas regiones de contactos compartidos. Por ejemplo, la región 3' de *Ibsp* es rica en interacciones con los tres promotores, es decir, podría tratarse de un MCR común para el locus SCPP ácido expresado en hueso. Otras dos regiones comunes se encuentran en los cuerpos génicos de *Mepe* y *ckd12*. Por otro lado, hay regiones que son comunes con dos promotores, como es el caso de las interacciones compartidas entre *Dmp1* e *Ibsp*, que incluye el contacto

entre ambos promotores del que se habló previamente, así como la interacción con la región promotora de *Odam*.

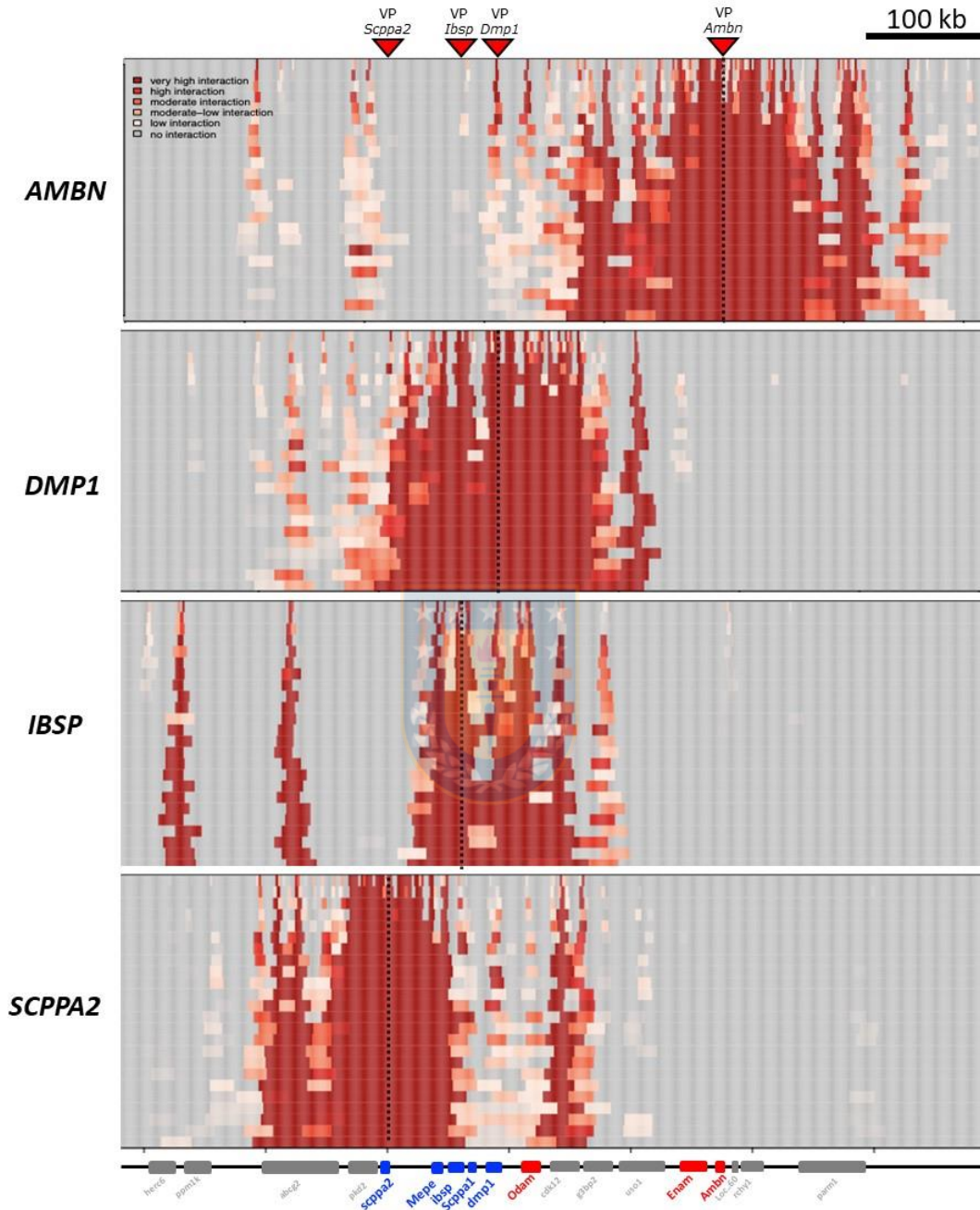


Figura 39: **Todos contra todos**

Se alinearon todas las regiones de interacción significativa a lo largo de una ventana ~1MB que incluye los *Vp* de los genes SCPP. Un gradiente que va del gris al rojo implica nula o alta interacción respectivamente. Cada línea negra señala los *Vp* para cada uno de los genes señalados arriba con un triángulo rojo invertido. Abajo las referencias de genes, con cajas azules (genes SCPP ácidos), rojos (SCPP ricos en P/Q) y grises (genes vecinos).

7. DISCUSIÓN

Resulta prácticamente intuitiva la idea de que genes que conservan sintenia puedan ser corregulados y/o co-expresados en los procesos del desarrollo de un organismo. Los genes *Hox*, *Six* e *Irx* son ejemplos claros de este tipo de eventos biológicos (Gomez-Marin et al., 2015; Noordermeer et al., 2011; Tena et al., 2011). A su vez, parece bastante razonable que, si estos genes no se han “atomizado” en el genoma desde su origen, es porque regiones reguladoras no codificantes pueden constreñir su ubicación cromosomal a lo largo de millones de años, por lo que una afectación de estos elementos podría causar estragos en la supervivencia de la especie. No obstante, lo aparentemente intuitivo debe ser también demostrado. Es así como el objetivo principal de este trabajo fue determinar si la transcripción de genes del locus SCPP involucra la acción de múltiples elementos reguladores que contactan simultáneamente y corregulan miembros coexpresados en etapas específicas del proceso de diferenciación del linaje osteogénico en *X. tropicalis*.

La etapa específica donde se centró este trabajo fue en osteoblastos, principalmente porque los ensayos reporteros pierden su relevancia biológica si no están realizados en el contexto celular adecuado. Nuestro laboratorio ya había estandarizado el cultivo primario de este tipo celular a partir de tejido óseo de renacuajos de *X. tropicalis* (Bertin et al., 2015), de tal manera que fue factible obtener la gran cantidad de células requeridas para realizar la técnica de 4C. Además, contamos con datos de RNA-seq y ChIP-seq de tejido óseo y osteoblastos respectivamente, lo que posibilitaba indagar el posible paisaje regulatorio asociado al locus SCPP.

Hasta el momento se conoce relativamente bien la arquitectura regulatoria del genoma de mamíferos, principalmente en células embrionarias de ratón y en líneas celulares humanas (Dixon et al., 2012; Schoenfelder & Fraser, 2019). No obstante, los estudios de arquitectura regulatoria en *X. tropicalis* son prácticamente nulos, con la excepción del reciente estudio de (Niu et al., 2020) . Particularmente, no tenemos referencias que sustenten cómo es la arquitectura de la cromatina en células óseas.

La mayor parte de estudios de arquitectura que se han hecho hasta el momento son de genes canónicos que tienen sintenia conservada a través de los metazoos. Los GRBs de estos genes han permanecido por más de 550 ma (*Hox*, *Irx*, *Hedgehog*, *Pax* y *FoxA*). Este trabajo tiene la relevancia de que por primera vez se estudia la arquitectura regulatoria de genes que no son factores de transcripción, sino que son elementos de matriz celular sumamente relevantes para la mineralización de tejidos esqueléticos y que están asociados con la evolución de dichos tejidos

desde su surgimiento hace ~450 ma, justo después de la segunda ronda de duplicaciones del genoma completo. Además, procesos de nuevas duplicaciones locales originaron nuevos genes en el locus con una importante diversificación de funciones en mamíferos sin modificar radicalmente su posicionamiento cromosomal. Por lo tanto, en este trabajo fue fundamental estudiar la arquitectura regulatoria del locus SCPP en un clado representante de los vertebrados no amniotas.

7.1 Expresión de los genes SCPP y SPARC (Una actualización)

Los genes SCPP ácidos codifican para una familia de proteínas que son encontradas en tejidos mineralizados y en fluidos tisulares como la leche, lágrimas y saliva. Sus funciones están relacionadas principalmente con la unión a calcio (Hunter & Goldberg, 1993; Kawasaki, 2009). Estructuralmente, los genes tienen características que permiten agruparlos y las proteínas resultantes poseen dominios funcionales que las relacionan evolutivamente. Estas proteínas derivaron de un ancestro común y están principalmente involucradas en la regulación de la mineralización de hueso y dentina (Kawasaki et al., 2004).

La osteopontina, codificada por el gen *Spp1*, fue detectada por inmunohistoquímica en matriz extracelular de células óseas de mamíferos (Mark et al., 1987). También fue detectada por hibridación *in situ* en odontoblastos y condrocitos hipertróficos de modelos murinos (Nomura et al., 1988). El gen *Scppa2* descrito exclusivamente en el genoma de *X. tropicalis*, es el más cercano en cuanto a estructura, función y posicionamiento en el genoma de *Spp1* de mamíferos. En *X. tropicalis*, fue observado en osteoblastos, osteocitos de hueso mandibular y odontoblastos (Kawasaki, 2009). Nuestros resultados concuerdan con lo observado anteriormente en hueso de *X. tropicalis* e hicimos la observación que *Scppa2* sólo se expresa en algunos osteoblastos, posiblemente células activas en el proceso de secreción de la matriz y sorpresivamente, también lo vimos fuertemente en cartílago mineralizado de extremidades y vertebras, algo que hasta el momento no estaba descrito.

IBSP, actúa como un potente nucleador de la mineralización. La proteína tiene propiedades bioquímicas que le permiten interactuar con la hidroxiapatita (Vincent & Durrant, 2013). Está presente en osteoblastos maduros y en osteocitos de modelo murino (Chen et al., 1991). En *X. tropicalis* la expresión de este gen fue observada en hueso de mandíbula inferior de rana adulta aunque no se indican los niveles de expresión en osteoblastos respecto a osteocitos (Kawasaki, 2009). Nuestro laboratorio describió la expresión de este gen principalmente en osteoblastos de extremidad y sólo parcialmente en osteocitos (Espinoza et al., 2010). En este

trabajo, un análisis con sondas intrónicas nos permitió sugerir que *lbsp* se expresa exclusivamente en osteoblastos y que la débil señal observada en osteocitos se debe a la presencia de transcritos que permanecieron estables en el citoplasma después de la diferenciación.

El gen *Dmp1* codifica para una proteína secretada en la formación de hueso y dentina. Su expresión ha sido identificada parcialmente en osteoblastos que secretan el osteoide y fuertemente en osteocitos (Kamiya & Takagi, 2001; Silvent et al., 2013). En *X. tropicalis*, *Dmp1* fue observado en osteocitos (Espinoza et al., 2010) y muy ocasionalmente en la zona perinuclear de algunos osteoblastos que probablemente están en proceso de diferenciación osteocítica. Proponemos que las células que se encuentran exactamente al inicio de la transición de osteoblasto a osteocito son muy escasas, y que por este motivo en este trabajo no observamos expresión de DMP1 en osteoblastos y sólo lo observamos en osteocitos, incluso en aquellos que apenas quedaron atrapados en su matriz o recién diferenciados.

Respecto a los genes *Scppa1* y *Mepe*, el primero sólo ha sido identificado en *X. tropicalis* por RT-PCR expresado en diente pero no en hueso (Kawasaki, 2009) mientras que el segundo no ha sido observado en hueso de *X. tropicalis*. Los transcritos de *Mepe* han sido identificados en osteoblastos y principalmente en osteocitos de modelo murino y en humano se expresa exclusivamente en osteocitos. Los datos de RNA-seq de nuestro laboratorio no muestran expresión de *Mepe* en hueso, por lo que la expresión de este gen es divergente respecto a lo que sucede en mamíferos.

Con este trabajo demostramos los patrones de expresión específicos de los genes SCPP acídicos en células de hueso de *X. tropicalis*. Los transcritos de *lbsp* fueron observados exclusivamente en osteoblastos mientras que *Dmp1* exclusivamente en osteocitos. *Scppa2* fue observado en osteoblastos, osteocitos y también en cartílago calcificado de extremidad de renacuajos. Aunque no es parte de este trabajo, es importante mencionar que *Scppa2* también fue confirmado en cartílago mineralizado de vertebras. *Mepe* y *Scppa1* no fueron evaluados por *His* porque no se expresan en hueso (datos de RNA-seq).

7.2 Arquitectura regulatoria del locus SCPP

Con este trabajo demostramos por ensayos con reporteros al menos 1 enhancer proximal B que activa fuertemente el promotor de *lbsp* y un enhancer distal A que lo activa débilmente. Sorpresivamente, dos regiones genómicas enriquecidas con marcas de histonas asociadas a enhancer activos irrumpieron negativamente la expresión del gen *Scppa2* en osteoblastos, por

lo que puede tratarse de elementos que reprimen la actividad transcripcional. Estas mismas regiones no alteraron la actividad de los promotores de *lbsp* y *Dmp1*. Por el contrario, el enhancer A, que fue parcialmente activo en *lbsp*, tuvo actividad represora sobre el promotor de *Scppa2* (Figura 41). Esta represión puede ser explicada por la incompatibilidad entre enhancers y promotores, por ejemplo, el enhancer A que activa parcialmente *lbsp* tiene una actividad inversa en el promotor de *Scppa2*. Una segunda explicación podría involucrar marcas de histonas no canónicas que probablemente utilicen algunos enhancers activos específicamente en *X. tropicalis*. No fue posible probar el enhancer B en *Scppa2*, por lo que se sugiere como proyección.

En cuanto a la arquitectura regulatoria, demostramos que hay una riqueza de interacciones entre los promotores de los genes SCPP principalmente con blancos de genes del mismo locus y parcialmente con otras regiones distales. También que los tres promotores tienen contactos específicos, algunos proximales y otros distales, incluso a regiones >100 kb. Un hallazgo relevante fue la interacción espacial entre los promotores de *lbsp* y *Dmp1*, así como un fuerte contacto entre *lbsp* y el enhancer proximal B, el cual tiene marca cercana de cromatina abierta (ATAC), conservación genómica con otros vertebrados y mostró actividad en ensayos reporteros (Figura 41). Desde el punto de vista de la regulación de la transcripción, esta conformación sugiere un micro-dominio de asociación de promotores SCPP que podría ser el mecanismo que explique y posibilite el *switch* regulatorio de la transición osteoblasto-osteocito.

7.3 Dos promotores que interactúan se expresan diferencialmente

Aunque la literatura ha mostrado que los promotores pueden actuar como enhancers de otros genes (Barutcu et al., 2014), llama la atención que dos promotores que se expresan de forma diferencial (*lbsp* en osteoblastos y *Dmp1* en osteocitos) estén interactuando. Los datos de ChiP-seq muestran que la región promotora de *Dmp1* no tiene marcas de histonas asociadas a enhancers putativos ni regiones de cromatina abierta (Figuras 32F y 33), entonces, surge la pregunta ¿cuál sería la relevancia funcional de la interacción entre los promotores *lbsp-Dmp1* en el contexto de la regulación transcripcional en hueso? Un posible modelo podría sugerir un micro-dominio. En dicha estructura, el promotor de *lbsp* está activo y *Dmp1* reprimido. Tan pronto los factores osteocíticos se activan, la cromatina reprimida, con las interacciones preestablecidas, recibe factores de transcripción que rápidamente activan los enhancers de *Dmp1* explicando el *switch* regulatorio que lleva a la transcripción de *Dmp1* en un contexto osteocítico (Figura 40).

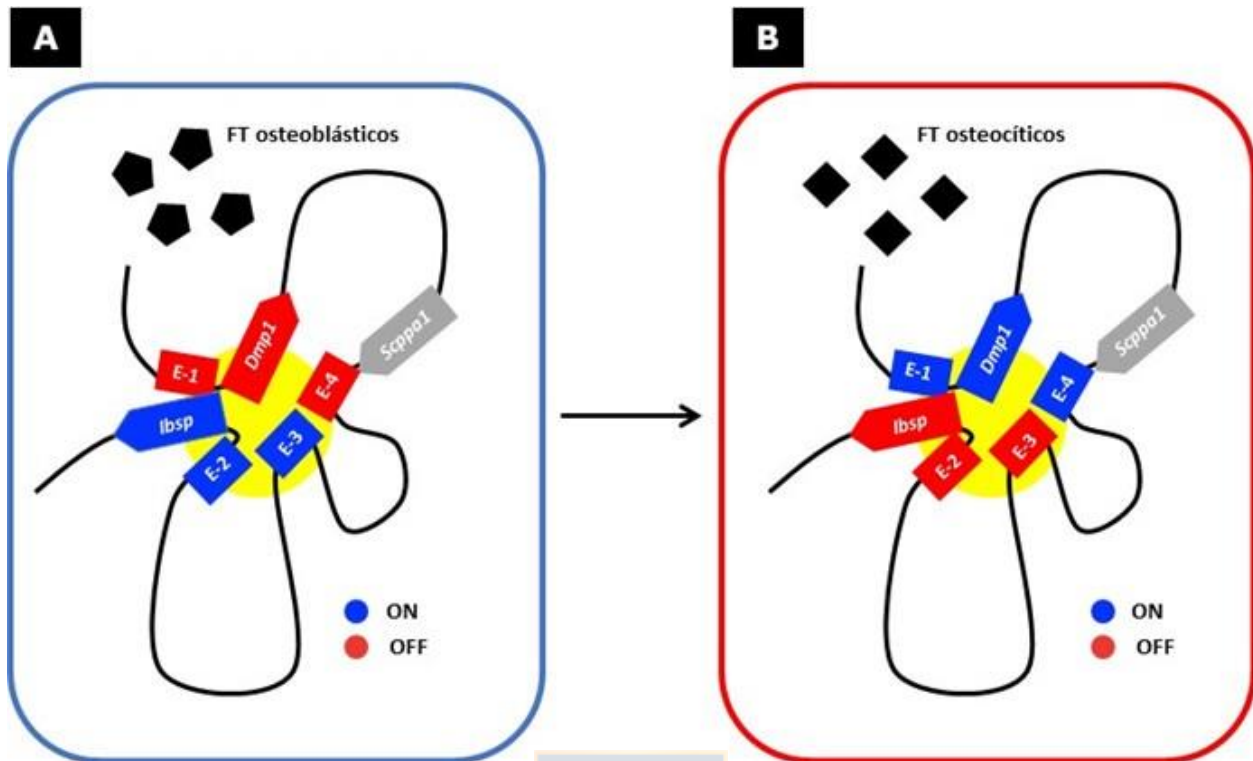


Figura 40: **Modelo de interacción *Ibsp-Dmp1***

En este modelo los promotores de ambos genes (*Ibsp-Dmp1*) se contactan en el contexto osteogénico. Todos los enhancers (1-4) (esquematisados en función de las regiones de interacción significativas para ambos promotores) están contactando a sus genes blanco en un microdominio (circulo amarillo). En A, la actividad del gen *Ibsp* va a depender de los FT específicos de osteoblastos y del promotor activo, donde se ensambla la maquinaria transcripcional. En B, cuando estos factores se vuelven osteocíticos, los enhancers de *Dmp1* y el promotor se activan. La arquitectura de la cromatina de un microdominio favorece un núcleo transcripcional que explica el *switch* regulatorio de la transición osteoblasto-osteocito. En este contexto, probablemente enhancers distales dinámicos o específicos de tipo celular contactan diferencialmente a cada promotor incrementando la transcripción de manera temporal-específica. En este modelo el extremo 3' de *Scppa1* (que se expresa en diente) tiene una región (E-4) que interactúa con ambos promotores.

Cruzando los datos de *ChIP-seq* con 4C observamos que los enhancers funcionales de *Ibsp* y *Sparc* se encuentran en regiones ricas de interacciones. Asociadas a estas regiones se encontró también enriquecimiento de secciones de cromatina abierta (Figura 41A). Otras interacciones fueron relevantes porque presentan frecuencias que colocalizan para los tres promotores del locus SCPP (*Dmp1*, *Ibsp* y *Scppa2*). Al menos dos regiones del locus son

contactadas significativamente por los tres promotores y otras por al menos dos promotores, por ejemplo, los contactos entre *lbsp* y *Dmp1* (Figura 40). A grandes rasgos podemos indicar que el gen divergente en cuanto a la colocación de contactos fue *Scppa2*. Algo que es relevante de este gen es que además de expresarse en diente y hueso (Kawasaki, 2009), se expresa en cartílago mineralizado (datos de este trabajo) y riñón (datos RNA-seq, colaboración con Francia) (Figura 41A).

Como control se realizó también el 4C de *Ambn*, un gen expresado principalmente en diente, pero nulo en hueso. Las interacciones de *Ambn* en el contexto de osteoblastos fueron divergentes respecto a los genes acídicos, y sólo un par de interacciones son compartidas con *Dmp1*, sugiriendo que los SCPP acídicos se encuentran en un subdominio independiente de los SCPP ricos en P/Q.

7.4 Arquitectura regulatoria de *Sparc*

Se demostró la actividad funcional de dos enhancers (B y C) que activan el promotor. Uno de estos dos enhancers, es proximal a *Sparc* y es el que presentó mayor actividad. Nuestros datos de 4C sugieren que *Sparc* interactúa principalmente con elementos proximales y significativamente con al menos tres regiones distales río arriba del gen *Fat2* (Figura 41B).

Dos regiones ricas de interacción fueron demostradas alrededor del promotor. Una que abarca hasta 12 kb río arriba y que coincide en parte con los dos MCRs probados en ensayos reporteros y otra a lo largo del primer intrón. Como perspectiva, esta última región debería ser explorada funcionalmente ya que tiene 3 elementos conservados (Figura 41B).

Resultados de experimentos de HiC y MicroC realizados con diferentes células humanas y posicionados en la región de *Sparc* muestran una riqueza de contactos río arriba del promotor en lo que se podría delimitar como un microdominio. Resultados de MicroC en fibroblastos permiten observar, a una excelente resolución (2 Kb), los contactos específicos en la región río arriba de *Sparc* (Anexo 4). La proyección de puntos esquinados (Figura 43 en anexos) permite observar un gran punto (flecha negra) que demarca los límites del TAD. En la imagen se observan cada uno de los bystander ATOX1 y FAT2 justo en la periferia del TAD, pero en diferentes microdominios. Además, un subTAD, señalado con un triángulo (Figura 43 en anexos) separa a ATOX1 pero deja interactuando a FAT2 con SPARC. Esta misma observación fue hecha en *X. tropicalis* a partir de los resultados de 4C, donde los contactos distales que se encuentran río arriba de SPARC contactan al promotor. Esto podría considerarse otro ejemplo de genes que se quedaron unidos en la evolución ya que en *X. tropicalis* *Sparc* se encuentra en sintenia con los

genes *Atox1* y *Fat2*, además, dentro de los intrones de estos genes o en regiones intergénicas proximales a ellos hay algunas regiones que contactan al promotor de *Sparc*, lo que podría explicar también la conservación de sintenia de *Sparc-Fat2* en la evolución de vertebrados.

Los TADs pueden definirse funcionalmente como andamios dentro de los cuáles hay mayor probabilidad de interacción, pero donde sólo algunos contactos son tejido-específicos. Estos contactos dinámicos sólo fueron observados una vez que se incrementó la resolución de 1-2 kb a través de la técnica de MicroC. Nuestro trabajo tiene la gran ventaja que usamos enzimas que cortan con alta frecuencia, por lo que logramos alcanzar una resolución similar con la única desventaja de que sólo se pueden estudiar contactos específicos de los promotores seleccionados.



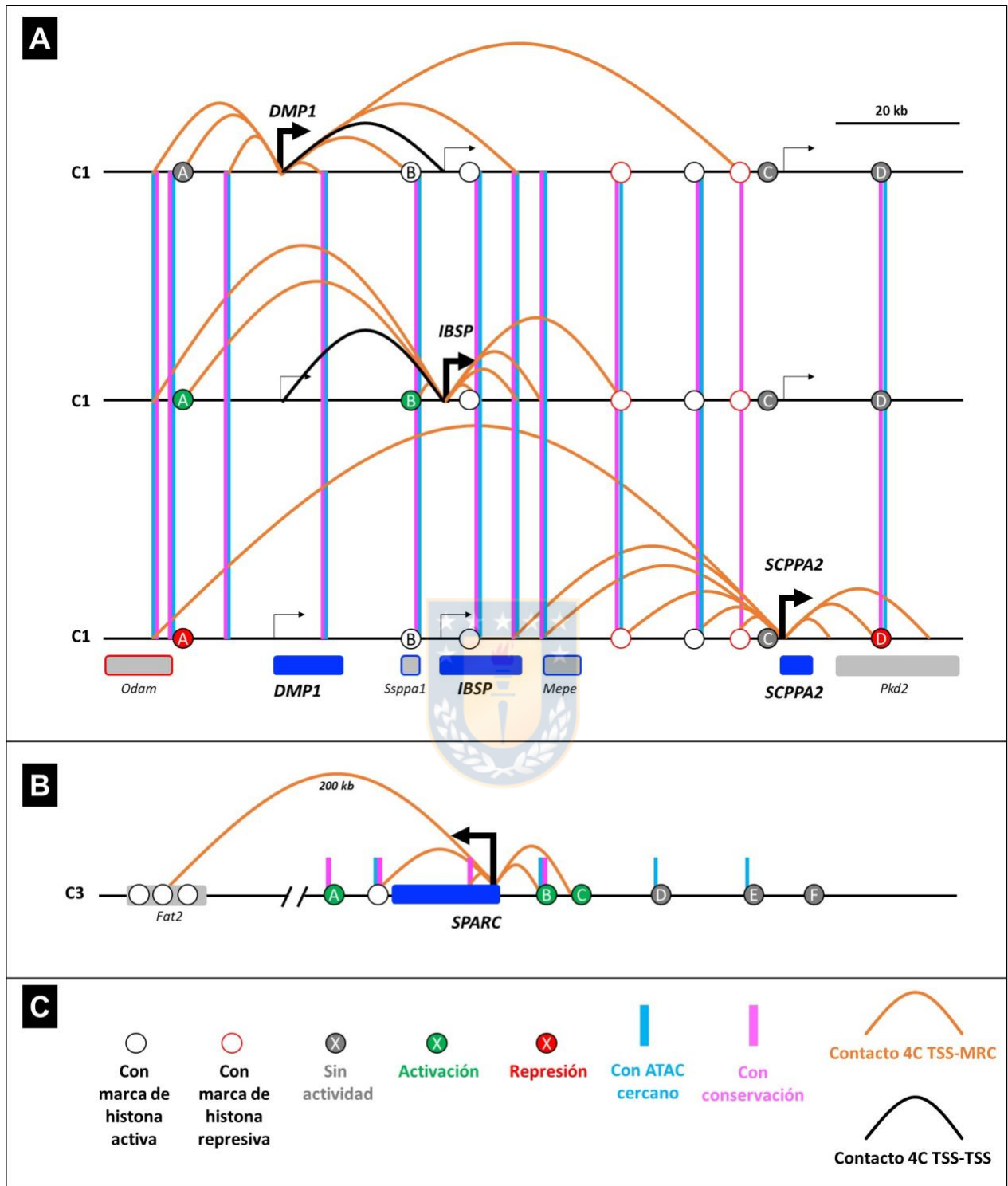


Figura 41: Resumen de hallazgos en el locus SCPP

En A, se muestra una ventana de 140 Kb (a escala) del cromosoma 1 (C1) que abarca el locus SCPP. Cada ovalo representa los diferentes MCRs ligados al locus. Para *Dmp1* se probaron 3 MCRs en ensayos reporteros (A, C y D) que no mostraron actividad. Para *Ibsp* se probaron 4 MCRs (A-D), A y B presentaron

actividad mientras que C y D fueron inactivos. En *Scppa2* los MCRs A y D tienen actividad represiva (óvalos rojos), C fue inactivo (ovalo gris) y B no fue probado en este estudio (ovalo blanco). Cada fecha negra representa el TSS de cada uno de los genes y abajo se muestra la anotación correspondiente (azul = genes SCPP ácidos y en gris los elementos bystander). Las cajas horizontales (rosas) representan regiones conservadas con otros vertebrados y las azules marcas de ATAC (cromatina abierta). Los óvalos blancos vacíos muestran regiones de marcas de cromatina asociados a MCRs putativos activos mientras que los óvalos de línea roja vacíos son MCRs con marcas de histonas represivas. Las líneas naranjas muestran las regiones significativas de interacción (4C) entre promotores y regiones del genoma ligadas al locus SCPP y las líneas negras los contactos entre los promotores *Dmp1-lbsp*. En B, se muestra el locus de SPARC que se encuentra en el cromosoma 3, al menos 3 MCRs mostraron actividad en ensayos reporteros. El MCR B muestra una rica región de contactos 4C además de conservación y cromatina abierta. Una región rica en contactos se encuentra a 200 kb coincidiendo justamente con la riqueza de contactos en mamíferos en la región que abarca el gen *FAT2* (datos de HiC en fibroblastos de humanos). En C, los símbolos con leyendas asociadas a cada uno de los paneles A y B.

7.5 CTCF en la arquitectura regulatoria del locus SCPP

Tal como fue expuesto en la introducción, una de las proteínas maestras de arquitectura de la cromatina es el factor de unión a CCCTC o CTCF. Esta proteína puede interactuar con otras proteínas CTCF así como con la cohesina generando un bucle que acerca espacialmente regiones alejadas linealmente en el genoma (Hansen, 2020).

Mediante estudios de *ChIP-seq*, el factor CTCF se ha encontrado enriquecido en los límites entre TADs y cuando es alterado, los promotores de genes pueden cambiar sus frecuencias de interacción con diferentes enhancers llevando a un desorden en la expresión génica. Debido a que los estudios de arquitectura regulatoria en *X. tropicalis* son prácticamente nulos, no se tiene acceso a datos en donde se puedan indagar posibles proteínas que participen en la arquitectura del locus SCPP. Al no contar con *ChIP-seq* de CTCF en hueso de *X. tropicalis*, se recurrió a hacer, de forma preliminar, una búsqueda bioinformática de sitios de unión putativos a CTCF.

Se encontraron al menos 4 sitios significativos CTCF convergentes, es decir, asociados a formación de bucles de cromatina (Figura 41). También se encontraron sitios de unión a CTCF internos que aunque no fueron convergentes pueden tratarse de elementos insuladores. Regiones ricas de unión a CTCF se ubican en dos extremos del locus SCPP, a una distancia ~600 kbp, lo que sugiere, en primera instancia, que el locus SCPP en *X. tropicalis* se encuentra en un Dominio de Asociación Topológica (TAD). Otros sitios de unión a CTCF separan los genes del locus SCPP ácidos expresados de los genes ricos en P/Q, lo que sugiere que ambos grupos se encuentran en dos subTADs. Por lo tanto, estas frecuencias de interacción significativas en el locus SCPP ácido sugieren una función de la regulación génica derivada de la arquitectura de

la cromatina. Algo muy similar sucede en el genoma humano donde los SCPP ácidos se encuentran en un TAD, pero dentro del TAD se pueden encontrar subTADs e incluso microdominios, por ejemplo, en células humanas, hay claramente dos microdominios que separan *Ibsp-Dmp1-Mepe* de *Spp1* (Anexo 3).

Sería fundamental, a manera de proyección, realizar el ChIP-seq de CTCF específicamente en osteoblastos de *X. tropicalis* para cruzar con los datos de 4C y confirmar esta sugerencia. No obstante, de manera teórica, se propone el posible modelo de arquitectura reguladora sugerido por los sitios de unión a CTCF encontrados en el locus SCPP (Figura 40). Observe como en este modelo los genes SCPP interactúan principalmente en un subTAD, mientras que un posible insulador (río arriba de ODAM) separaría algunos contactos en otro subdominio que incluye los genes SCPP ácidos que aun comparte algunas interacciones con el subdominio de los genes ricos en P/Q. Esta conformación permite que los promotores de *Ibsp* *Dmp1* y *Scppa2* se mantengan en un mismo dominio de interacción.

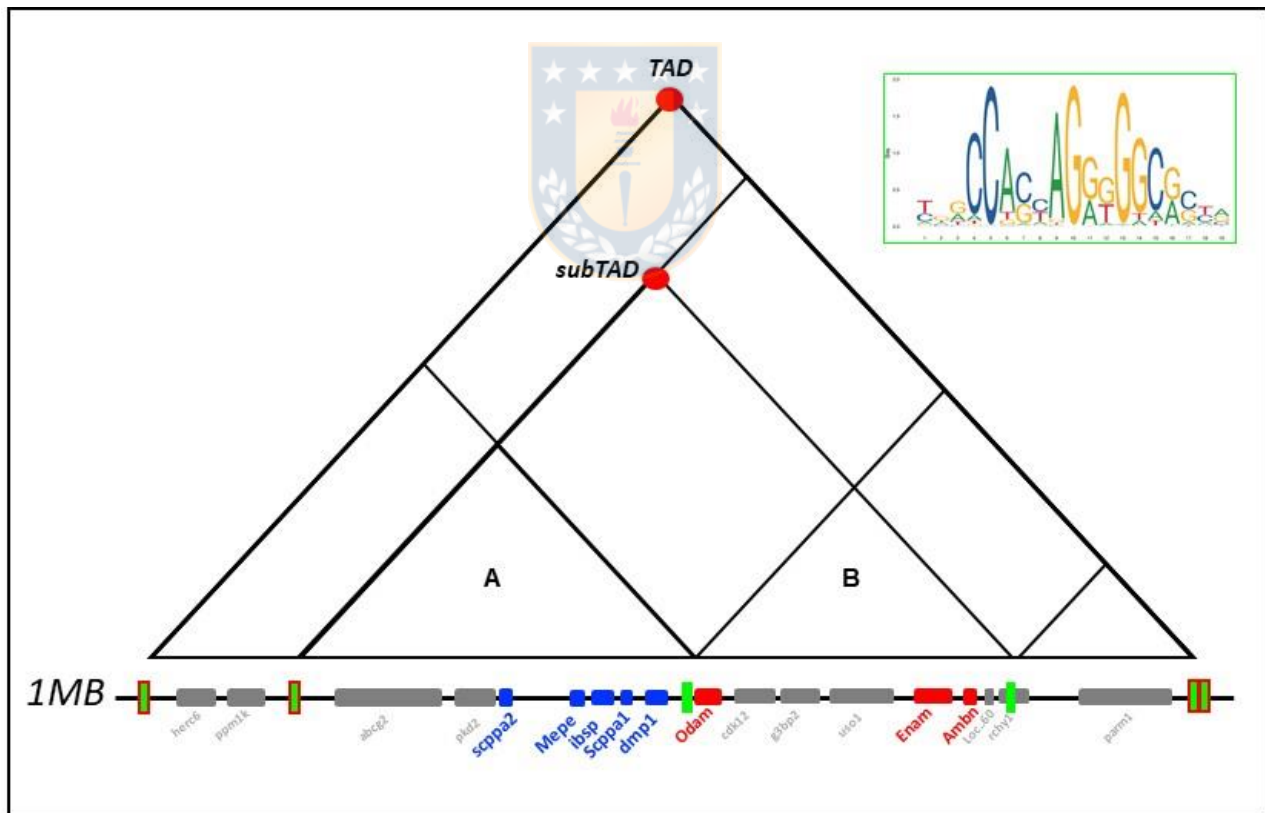


Figura 42: **Arquitectura regulatoria del locus SCPP**

Se realizó un modelo teórico basado en el enriquecimiento de motivos de unión para CTCF el cual es mostrado en el rectángulo verde. En la región que contiene el locus SCPP se encontraron al menos 6 sitios

de unión a CTCF de mayor significancia estadística ($p\text{-value} < 0.0000001$) analizadas en la plataforma The Meme Suite: Motif-based sequence analysis tools. Estas regiones son mostradas con rectángulos verdes a lo largo del 1 MB en la zona del cromosoma 1 que incluye el locus SCPP. Dos regiones distales en las afueras del locus muestran sitios de unión a CTCF convergentes, lo que sugiere la formación de un Dominio de Asociación Topológica. Dentro de este TAD, se observa un *subtad* con al menos dos subdominios (I y II) separados por un posible insulador río arriba de ODAM que divide las regiones de genes SCPP acídicos de los ricos en P/Q.

7.6 Evolución del locus SCPP y SPARC

El orden de los genes en los cromosomas de los metazoos ha sugerido una distribución al azar y no funcional, sin embargo, algunos estudios han demostrado que hay genes que se han mantenido unidos en pares o en cluster después de millones de años de evolución, sugiriendo funcionalidad de esta organización. Al parecer, estas asociaciones pueden permanecer en la evolución de genomas porque hay elementos *cis*-regulatorios compartidos que constriñen su evolución (Irimia et al., 2013; Kikuta et al., 2007). Ya sea que múltiples genes compartan elementos regulatorios transcripcionales o que estos elementos regulatorios se encuentran en zonas de genes vecinos “bystander” que los mantienen unidos a pesar de ser genes con funciones diferentes. Este tipo de organizaciones fueron denominadas bloques de regulación genómica (GRBs) y en los últimos años se demostró que suelen encontrarse dentro de dominios de asociación topológica o TADs (Harmston et al., 2017).

En este trabajo estudiamos el locus SCPP y *Sparc* y demostramos que dentro del locus SCPP existen interacciones que de ser funcionales serían las que mantienen los miembros del locus unidos a lo largo de millones de años de evolución de tetrápodos no amniotas. Muy probablemente, antes de las duplicaciones en tándem, un gen fundador contenía MCRs proximales que dirigían su expresión. Una situación hipotética post-duplicación dejaría unidos los genes recién duplicados, conservando los contactos preestablecidos del gen fundador y generando a su vez nuevos contactos, esto explicaría las interacciones observadas entre los promotores de *Dmp1* con *lbsp*, así como las ricas frecuencias de interacción en el micro-dominio del locus (Figura 43).

De forma relevante, *Sparc* también muestra una zona de interacciones alrededor del promotor, algo que puede ser asociado a un microdominio y que además se conserva en la estructura topológica en humano. Respecto a humano, existe un contacto con el bystander *FAT2* (Figura 43 en anexos) que genera un subTAD que se observó también en *X. tropicalis* con una región de interacción distal con *Fat2* (Figura 38). Aunque en humanos se muestra el resultado

derivado de células inmortalizadas de fibroblastos, en general esta estructura es conservada y puede ser a grandes rasgos comparada con los resultados de 4C obtenidos en *X. tropicalis*. Como proyección sería indispensable estudiar funcionalmente el contacto distal (Fat2)-Sparc de humano y de *X. tropicalis* para evaluar tanto su actividad sobre el promotor de *Sparc* como su conservación.

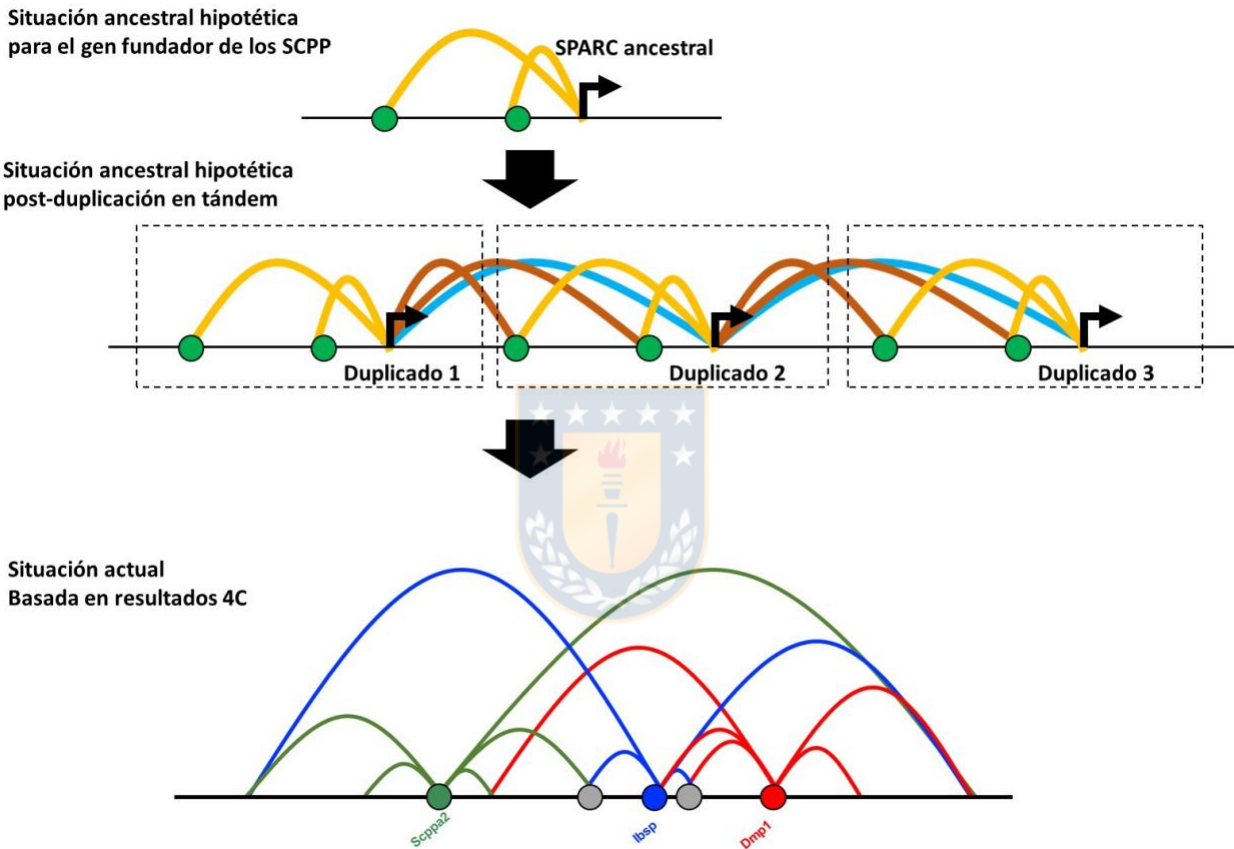


Figure 43: **Situación ancestral VS situación actual de los genes SCPP**

Un gen fundador ancestral probablemente era regulado por enhancers proximales. Posterior a la ronda de duplicaciones del genoma completo el gen Sparc1 duplicó en tándem. Una situación ancestral hipotética muestra que después de la duplicación los promotores de los genes SCPP ancestrales quedaron ligados. La situación actual generada por datos de 4C demuestra que los SCPP ácidos quedaron unidos compartiendo elementos reguladores, lo que explicaría el hecho de que los miembros del locus no se atomizaron en el genoma durante la evolución de vertebrados.

Finalmente, se propone un modelo basado en los resultados de las interacciones significativas mostradas en los dominiogramas de todos los promotores del locus SCPP (Figura 42). Al menos dos regiones son contactadas por los tres promotores: en *Ckd12* y en el extremo terminal de *Ibsp*. La mayor parte de regiones significativas son contactadas por al menos dos promotores, particularmente las asociadas a los promotores de *Ibsp* y *Dmp1*. De forma relevante, los genes del locus SCPP acídicos no expresados en hueso (*Mepe* y *Scppa1*) son ricos en frecuencias de contactos con al menos dos de los tres promotores. *Dmp1* y *Ambn* tienen interacciones en la región de los bystander *g3bp2* y *uso1*, lo que implica que ambos promotores contactan en regiones que se entrecruzan.

En conclusión, hay algunos contactos que son compartidos, es decir, una única región puede ser contada por más de 1 promotor. Por lo tanto, se sugiere un modelo donde la riqueza de contactos compartidos en el locus SCPP genera una arquitectura regulatoria que mantiene unidos los miembros del clúster, lo que explicaría por qué estos genes no se “atomizaron” en el genoma de los vertebrados.

Como proyección se sugiere evaluar funcionalmente las regiones ricas de interacción que estén asociadas, en orden de importancia a marcas de cromatina proximales, cromatina abierta, conservación con otros genomas de vertebrados e interacción significativa.

Todos los datos mostrados en este trabajo apuntaron a contestar la pregunta de investigación que inspiró esta tesis ¿se caracterizará el locus SCPP por la presencia de múltiples MCRs que contactan y corregulan genes coexpresados en la diferenciación del linaje osteogénico? Aunque la respuesta sigue en proceso: estamos realizando la réplica del 4C que validaría de mejor forma nuestros resultados y falta probar funcionalmente las nuevas regiones descubiertas por el ensayo de conformación de cromosoma, se puede indicar que los datos preliminares apuntan a una arquitectura regulatoria que se ha mantenido a lo largo de millones de años por la restricción evolutiva de separar elementos reguladores dentro del locus SCPP y *Sparc*. Finalmente, la activación transcripcional del locus SCPP involucra la acción de múltiples MCRs que contactan y coregulan promotores de miembros expresados en etapas específicas del proceso de diferenciación del linaje osteogénico. Probar funcionalmente las nuevas regiones confirmadas por una réplica de 4C posibilitaría en gran medida la corroboración de esta hipótesis.

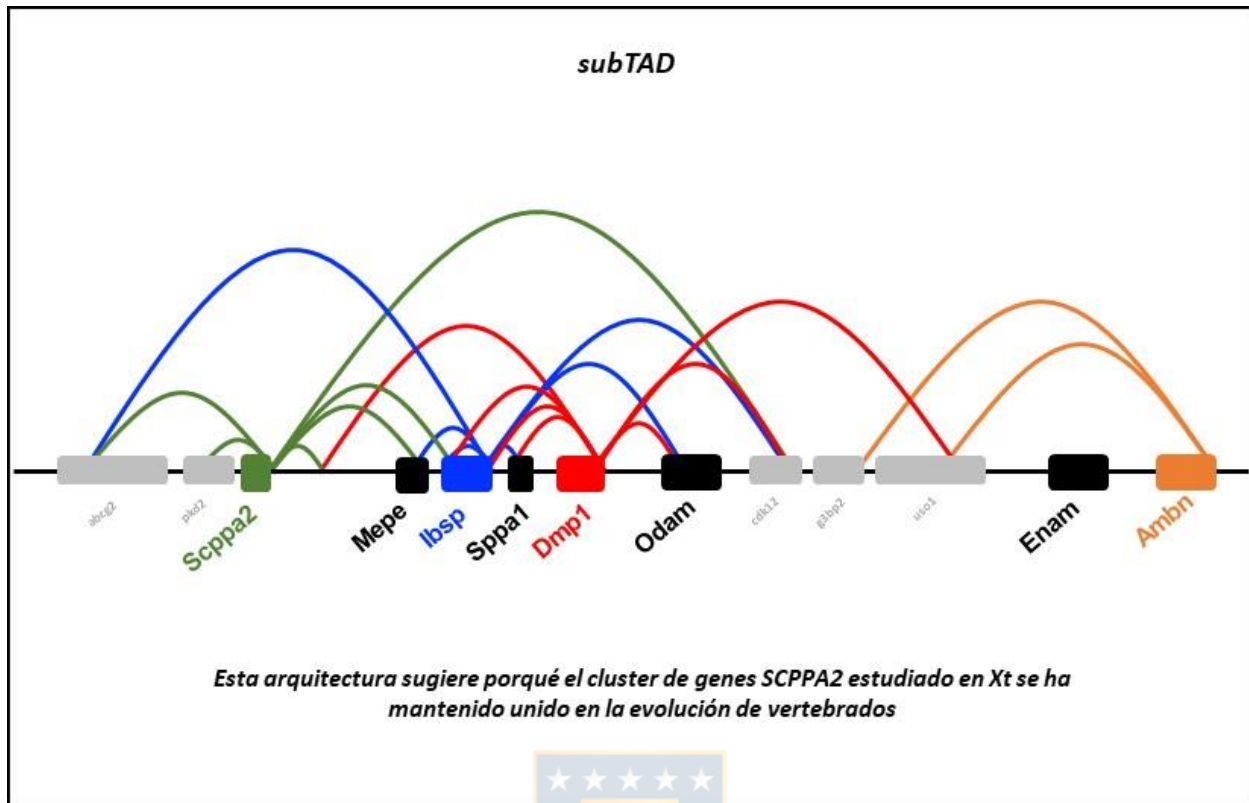


Figura 44: ¿Por qué los genes SCPP no se separaron en la evolución?

Tomando en cuenta los dominiogramas mostrados en la figura 40 se esquematizó un modelo de interacciones que sugiere porqué en la evolución del genoma de *X. tropicalis* los miembros del locus SCPP no se han separado y se mantienen en clúster. La unión putativa de CTCF en el locus SCPP sugiere que estas riquezas de frecuencias de interacción caen dentro de un subTAD.

8. CONCLUSIONES

- El gen *Dmp1* se activa tan pronto los osteoblastos maduros diferencian a osteocitos. Los resultados de este trabajo apoyan la idea que en *X. tropicalis* los MCRs que controlan la expresión de *lbsp* están activos sólo en osteoblastos y la señal débil observada en algunos osteocitos es probablemente ARN mensajero maduro tardío. Considerando que osteocitos diferencian directamente de osteoblastos, la expresión de *lbsp* reveló un dramático switch regulatorio en la transición osteoblasto-osteocito. *lbsp* es abruptamente apagado mientras *Dmp1* es rápidamente activado. Los datos derivados de *Sparc* revelaron el mismo patrón de actividad de *lbsp*.
- A diferencia de *lbsp* y *Sparc*, *Scppa2* se expresó espacialmente en osteoblastos, osteocitos y condrocitos calcificados, por lo que probablemente su regulación transcripcional es debido a MCRs independientes de los que controlan la expresión de otros genes del locus. La arquitectura reguladora mostro también que los contactos con el promotor de *Scppa2* divergen relativamente de *lbsp* y *Dmp1*.
- Por ensayos reporteros se demostró que existen al menos dos enhancers que activan el promotor de *lbsp*. El enhancer A es proximal a *Dmp1* y el enhancer B, de mayor actividad, es proximal al promotor de *lbsp*. El enhancer B tiene marcas de cromatina abierta y contactos proximales de 4C.
- Respecto al promotor *Sparc*, se demostró que tres enhancers tuvieron actividad transcripcional. Dos de ellos (B y C) tienen actividad fuerte en cultivo primario de osteoblastos, aunque esta actividad puede variar en el enhancer C, el cual fue activo principalmente en células en pasaje 1. Por lo que se sugiere que la actividad transcripcional depende también, probablemente, del pasaje celular. La región que abarca los enhancers B y C es rica en contactos 4C.
- Las frecuencias de interacción derivadas de datos de 4C con diferentes promotores (*Dmp1*, *lbsp*, *Scppa2* y *Ambn*) son enriquecidas dentro del locus SCPP. Las regiones de mayor interacción se observaron en elementos proximales de cada uno de los genes. El promotor de *Dmp1* es rico en interacciones proximales e interactúa con el promotor *lbsp*, lo que fue comprobado también cuando se realizó el 4C usando como *viewpoint* el promotor de *lbsp*.
- Todos los datos confirman nuestra hipótesis inicial, aunque se sugiere una réplica del experimento de 4C e incrementar la N en ensayos reporteros.

9. BIBLIOGRAFÍA

- Acemel, R. D., Tena, J. J., Irastorza-Azcarate, I., Marlétaz, F., Gómez-Marín, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S. G., Aldea, D., & Aury, J.-M. (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nature Genetics*, *48*(3), 336-341.
- Addison, W. N., Masica, D. L., Gray, J. J., & McKee, M. D. (2010). Phosphorylation-dependent inhibition of mineralization by osteopontin ASARM peptides is regulated by PHEX cleavage. *Journal of Bone and Mineral Research*, *25*(4), 695-705.
- Amândio, A. R., Lopez-Delisle, L., Bolt, C. C., Mascrez, B., & Duboule, D. (2020). A complex regulatory landscape involved in the development of mammalian external genitals. *Elife*, *9*, e52962.
- Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., & Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*, *340*(6137).
- Baht, G. S., Hunter, G. K., & Goldberg, H. A. (2008). Bone sialoprotein–collagen interaction promotes hydroxyapatite nucleation. *Matrix Biology*, *27*(7), 600-608.
- Balic, A., & Mina, M. (2011). Identification of secretory odontoblasts using DMP1-GFP transgenic mice. *Bone*, *48*(4), 927-937.
- Barros, N. M., Hoac, B., Neves, R. L., Addison, W. N., Assis, D. M., Murshed, M., Carmona, A. K., & McKee, M. D. (2013). Proteolytic processing of osteopontin by PHEX and accumulation of osteopontin fragments in Hyp mouse bone, the murine model of X-linked hypophosphatemia. *Journal of Bone and Mineral Research*, *28*(3), 688-699.
- Barutcu, A. R., Tai, P. W., Wu, H., Gordon, J. A., Whitfield, T. W., Dobson, J. R., Imbalzano, A. N., Lian, J. B., van Wijnen, A. J., & Stein, J. L. (2014). The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter. *Nucleic acids research*, *42*(16), 10360-10372.
- Beagan, J. A., & Phillips-Cremins, J. E. (2020, Jan). On the existence and functionality of topologically associating domains. *Nat Genet*, *52*(1), 8-16. <https://doi.org/10.1038/s41588-019-0561-1>
- Bertin, A., Hanna, P., Otarola, G., Fritz, A., Henriquez, J. P., & Marcellini, S. (2015). Cellular and molecular characterization of a novel primary osteoblast culture from the vertebrate model organism *Xenopus tropicalis*. *Histochemistry and cell biology*, *143*(4), 431-442.
- Bertrand, S., Fuentealba, J., Aze, A., Hudson, C., Yasuo, H., Torrejon, M., Escriva, H., & Marcellini, S. (2013, Apr 22). A dynamic history of gene duplications and losses characterizes the evolution of the SPARC family in eumetazoans. *Proc Biol Sci*, *280*(1757), 20122963. <https://doi.org/10.1098/rspb.2012.2963>
- Bogdanović, O., Fernandez-Miñán, A., Tena, J. J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S. J., Veenstra, G. J. C., & Gómez-Skarmeta, J. L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome research*, *22*(10), 2043-2053.
- Bourque, G., Pevzner, P. A., & Tesler, G. (2004, Apr). Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res*, *14*(4), 507-516. <https://doi.org/10.1101/gr.1975204>

- Brettmann, E. A., Oh, I. Y., & de Guzman Strong, C. (2018). High-throughput identification of gene regulatory sequences using next-generation sequencing of circular chromosome conformation capture (4C-seq). *JoVE (Journal of Visualized Experiments)*(140), e58030.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engström, P. G., & Frith, M. C. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, 38(6), 626-635.
- Chen, J., Shapiro, H. S., Wrana, J. L., Reimers, S., Heersche, J. N., & Sodek, J. (1991). Localization of bone sialoprotein (BSP) expression to sites of mineralized tissue formation in fetal rat tissues by in situ hybridization. *Matrix*, 11(2), 133-143.
- Chinsembu, K. C. (2012). Teeth are bones: Signature genes and molecules that underwrite odontogenesis. *Journal of Medical Genetics and Genomics*, 4(2), 13-24.
- Dao, L. T., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., & Stephen, T. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics*, 49(7), 1073.
- De Wit, E., Bouwman, B. A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M. J., Krijger, P. H., Festuccia, N., Nora, E. P., & Welling, M. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, 501(7466), 227-231.
- Debiais-Thibaud, M., Simion, P., Ventéo, S., Muñoz, D., Marcellini, S., Mazan, S., & Haitina, T. (2019). Skeletal mineralization in association with type X collagen expression is an ancestral feature for jawed vertebrates. *Molecular Biology and Evolution*, 36(10), 2265-2276.
- Dehal, P., & Boore, J. L. (2005, Oct). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*, 3(10), e314. <https://doi.org/10.1371/journal.pbio.0030314>
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558), 1306-1311.
- Delany, A., Amling, M., Priemel, M., Howe, C., Baron, R., & Canalis, E. (2000). Osteopenia and decreased bone formation in osteonectin-deficient mice. *The Journal of clinical investigation*, 105(7), 915-923.
- Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., & Vingron, M. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nature Genetics*, 51(8), 1263-1271. <https://www.nature.com/articles/s41588-019-0466-z>
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012, Apr 11). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), 376-380. <https://doi.org/10.1038/nature11082>
- Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development*, 134(14), 2549-2560. <https://doi.org/10.1242/dev.001065>
- Duboule, D. (2007, Jul). The rise and fall of Hox gene clusters. *Development*, 134(14), 2549-2560. <https://doi.org/10.1242/dev.001065>
- Dubrulle, J., & Pourquié, O. (2004). fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature*, 427(6973), 419-422.

- Enault, S., Muñoz, D., Simion, P., Ventéo, S., Sire, J.-Y., Marcellini, S., & Debais-Thibaud, M. (2018). Evolution of dental tissue mineralization: an analysis of the jawed vertebrate SPARC and SPARC-L families. *BMC evolutionary biology*, *18*(1), 1-12.
- Enault, S., Muñoz, D. N., Silva, W. T., Borday-Birraux, V., Bonade, M., Oulion, S., Ventéo, S., Marcellini, S., & Debais-Thibaud, M. (2015). Molecular footprinting of skeletal tissues in the catshark *Scyliorhinus canicula* and the clawed frog *Xenopus tropicalis* identifies conserved and derived features of vertebrate calcification. *Frontiers in genetics*, *6*, 283.
- Espinoza, J., Sanchez, M., Sanchez, A., Hanna, P., Torrejon, M., Buisine, N., Sachs, L., & Marcellini, S. (2010, Nov). Two families of *Xenopus tropicalis* skeletal genes display well-conserved expression patterns with mammals in spite of their highly divergent regulatory regions. *Evol Dev*, *12*(6), 541-551. <https://doi.org/10.1111/j.1525-142X.2010.00440.x>
- Espinoza, J., Sanchez, M., Sanchez, A., Hanna, P., Torrejon, M., Buisine, N., Sachs, L., & Marcellini, S. (2010). Two families of *Xenopus tropicalis* skeletal genes display well-conserved expression patterns with mammals in spite of their highly divergent regulatory regions. *Evolution & development*, *12*(6), 541-551.
- Feng, J. Q., Ward, L. M., Liu, S., Lu, Y., Xie, Y., Yuan, B., Yu, X., Rauch, F., Davis, S. I., & Zhang, S. (2006). Loss of DMP1 causes rickets and osteomalacia and identifies a role for osteocytes in mineral metabolism. *Nature genetics*, *38*(11), 1310-1315.
- Feng, J. Q., Zhang, J., Dallas, S. L., Lu, Y., Chen, S., Tan, X., Owen, M., Harris, S. E., & Macdougall, M. (2002). Dentin matrix protein 1, a target molecule for Cbfa1 in bone, is a unique bone marker gene. *Journal of Bone and Mineral Research*, *17*(10), 1822-1831.
- Fisher, L., Torchia, D., Fohr, B., Young, M., & Fedarko, N. S. (2001). Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochemical and biophysical research communications*, *280*(2), 460-465.
- Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., & Chan, W.-L. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, *538*(7624), 265-269. <https://www.nature.com/articles/nature19800>
- Gehrig, J., Reischl, M., Kalmár, É., Ferg, M., Hadzhiev, Y., Zaucker, A., Song, C., Schindler, S., Liebel, U., & Müller, F. (2009). Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nature methods*, *6*(12), 911-916.
- Gehrke, A. R., Schneider, I., De La Calle-Mustienes, E., Tena, J. J., Gomez-Marin, C., Chandran, M., Nakamura, T., Braasch, I., Postlethwait, J. H., & Gómez-Skarmeta, J. L. (2015). Deep conservation of wrist and digit enhancers in fish. *Proceedings of the National Academy of Sciences*, *112*(3), 803-808.
- Giorgetti, L., & Heard, E. (2016). Closing the loop: 3C versus DNA FISH. *Genome biology*, *17*(1), 1-9.
- Gomez-Marin, C., Tena, J. J., Acemel, R. D., Lopez-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Viernas, E., Bovolenta, P., Nobrega, M. A., Carvajal, J., & Gomez-Skarmeta, J. L. (2015, Jun 16). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A*, *112*(24), 7542-7547. <https://doi.org/10.1073/pnas.1505463112>
- Göndör, A., Rougier, C., & Ohlsson, R. (2008). High-resolution circular chromosome conformation capture assay. *Nature protocols*, *3*(2), 303-313.
- Gopalakrishnan, R., Thomas, P. E., Benson, M. D., Wang, D., & Franceschi, R. T. (2003). A homeodomain protein binding element in the bone sialoprotein promoter is critical for tissue-specific expression in bone. *Connective Tissue Research*, *44*(1), 154-160.

- Gordon, J. A., Tye, C. E., Sampaio, A. V., Underhill, T. M., Hunter, G. K., & Goldberg, H. A. (2007). Bone sialoprotein expression enhances osteoblast differentiation and matrix mineralization in vitro. *Bone*, *41*(3), 462-473.
- Gröschel, S., Sanders, M. A., Hoogenboezem, R., de Wit, E., Bouwman, B. A., Erpelinck, C., van der Velden, V. H., Havermans, M., Avellino, R., & van Lom, K. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*, *157*(2), 369-381.
- Håkeliën, A. M., Bryne, J. C., Harstad, K. G., Lorenz, S., Paulsen, J., Sun, J., Mikkelsen, T. S., Myklebost, O., & Meza-Zepeda, L. A. (2014). The regulatory landscape of osteogenic differentiation. *Stem Cells*, *32*(10), 2780-2793.
- Hall, B. K. (2005). *Bones and cartilage: developmental and evolutionary skeletal biology*. Elsevier.
- Hanna, P. (2014). *Control transcripcional óseo y repeticiones genómicas en Xenopus tropicalis: un análisis por RNA-seq y ChIP-seq in vivo* Universidad de Concepción]. Dirección de Postgrado
- Hansen, A. S. (2020, Dec). CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus*, *11*(1), 132-148. <https://doi.org/10.1080/19491034.2020.1782024>
- Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merckenschlager, M., & Lenhard, B. (2017). Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nature communications*, *8*(1), 1-13.
- Hartmann, C. (2009, Oct). Transcriptional networks controlling skeletal development. *Curr Opin Genet Dev*, *19*(5), 437-443. [https://doi.org/S0959-437X\(09\)00141-5](https://doi.org/S0959-437X(09)00141-5) [pii] 10.1016/j.gde.2009.09.001
- Higashibata, Y., Sakuma, T., Kawahata, H., Fujihara, S., Moriyama, K., Okada, A., Yasui, T., Kohri, K., Kitamura, Y., & Nomura, S. (2004). Identification of promoter regions involved in cell-and developmental stage-specific osteopontin expression in bone, kidney, placenta, and mammary gland: an analysis of transgenic mice. *Journal of Bone and Mineral Research*, *19*(1), 78-88.
- Hoffman, M. M., Ernst, J., Wilder, S. P., Kundaje, A., Harris, R. S., Libbrecht, M., Giardine, B., Ellenbogen, P. M., Bilmes, J. A., & Birney, E. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic acids research*, *41*(2), 827-841.
- Hojo, H., & Ohba, S. (2020). Gene regulatory landscape in osteoblast differentiation. *Bone*, 115458.
- Holm, E., Aubin, J. E., Hunter, G. K., Beier, F., & Goldberg, H. A. (2015, Feb). Loss of bone sialoprotein leads to impaired endochondral bone development and mineralization. *Bone*, *71*, 145-154. <https://doi.org/10.1016/j.bone.2014.10.007>
- Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S. J., Bogdanovic, O., Lister, R., & Veenstra, G. J. C. (2015, Dec 18). Embryonic transcription is controlled by maternally defined chromatin state. *Nat Commun*, *6*, 10148. <https://doi.org/10.1038/ncomms10148>
- Hou, C., Li, L., Qin, Z. S., & Corces, V. G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular cell*, *48*(3), 471-484.
- Hsieh, T.-H. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Molecular cell*.
- Hunter, G. K. (2013, Oct). Role of osteopontin in modulation of hydroxyapatite formation. *Calcif Tissue Int*, *93*(4), 348-354. <https://doi.org/10.1007/s00223-013-9698-6>

Hunter, G. K., & Goldberg, H. A. (1993, Sep 15). Nucleation of hydroxyapatite by bone sialoprotein. *Proc Natl Acad Sci U S A*, 90(18), 8562-8565. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8397409

Ibrahim, D. M., & Mundlos, S. (2020, Mar 17). The role of 3D chromatin domains in gene regulation: a multi-faceted view on genome organization. *Curr Opin Genet Dev*, 61, 1-8. <https://doi.org/10.1016/j.gde.2020.02.015>

Inukai, S., Kock, K. H., & Bulyk, M. L. (2017, Apr). Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev*, 43, 110-119. <https://doi.org/10.1016/j.gde.2017.02.007>

Irimia, M., Maeso, I., Roy, S. W., & Fraser, H. B. (2013). Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends in Genetics*, 29(9), 521-528.

Irimia, M., Royo, J. L., Burguera, D., Maeso, I., Gómez-Skarmeta, J. L., & Garcia-Fernandez, J. (2012). Comparative genomics of the Hedgehog loci in chordates and the origins of Shh regulatory novelties. *Scientific reports*, 2(1), 1-9.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., & Levine, S. S. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), 430-435.

Kamiya, N., & Takagi, M. (2001). Differential expression of dentin matrix protein 1, type I collagen and osteocalcin genes in rat developing mandibular bone. *The Histochemical journal*, 33(9-10), 545-552.

Kawasaki, K. (2009, Mar). The SPP gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev Genes Evol*, 219(3), 147-157. <https://doi.org/10.1007/s00427-009-0276-x>

Kawasaki, K., & Amemiya, C. T. (2014, Sep). SPP genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. *J Exp Zool B Mol Dev Evol*, 322(6), 390-402. <https://doi.org/10.1002/jez.b.22546>

Kawasaki, K., Buchanan, A. V., & Weiss, K. M. (2007). Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs*, 186(1), 7-24.

Kawasaki, K., Suzuki, T., & Weiss, K. M. (2004, Aug 3). Genetic basis for the evolution of vertebrate mineralized tissue. *Proc Natl Acad Sci U S A*, 101(31), 11356-11361. <https://doi.org/10.1073/pnas.0404279101>

Kawasaki, K., & Weiss, K. M. (2006, May 15). Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Zool B Mol Dev Evol*, 306(3), 295-316. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16358265

Kawashima, N., & Okiji, T. (2016). Odontoblasts: Specialized hard-tissue-forming cells in the dentin-pulp complex. *Congenital anomalies*, 56(4), 144-153.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engström, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., & Howe, K. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome research*, 17(5), 545-555.

- Koehler, A., Desser, S., Chang, B., MacDonald, J., Tepass, U., & Ringuette, M. (2009). Molecular evolution of SPARC: absence of the acidic module and expression in the endoderm of the starlet sea anemone, *Nematostella vectensis*. *Development genes and evolution*, 219(9-10), 509-521.
- Komori, T. (2010). Regulation of bone development and extracellular matrix protein genes by RUNX2. *Cell and tissue research*, 339(1), 189.
- Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., Parsi, K. M., Yang, L., Maehr, R., & Mirny, L. A. (2020). Ultrastructural details of mammalian chromosome architecture. *Molecular cell*.
- Krijger, P. H., Geeven, G., Bianchi, V., Hilvering, C. R., & de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods*, 170, 17-32.
- Kuraku, S., & Meyer, A. (2009). The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int J Dev Biol*, 53(5-6), 765-773. <https://doi.org/10.1387/ijdb.072533km>
- Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics*, 12(14), 1725-1735.
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., & Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), 289-293.
- Lonfat, N., & Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS letters*, 589(20), 2869-2876.
- Long, F. (2011). Building strong bones: molecular regulation of the osteoblast lineage. *Nat Rev Mol Cell Biol*, 13(1), 27-38. <https://doi.org/nrm3254> [pii] 10.1038/nrm3254
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., & Mundlos, S. (2015, May 21). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), 1012-1025. <https://doi.org/10.1016/j.cell.2015.04.004>
- Lynch, M., & Conery, J. S. (2000, Nov 10). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), 1151-1155. <https://doi.org/10.1126/science.290.5494.1151>
- Maeso, I., Acemel, R. D., & Gómez-Skarmeta, J. L. (2017). Cis-regulatory landscapes in development and evolution. *Current opinion in genetics & development*, 43, 17-22.
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of genetics*, 92(1), 155-161.
- Maginot, M., Lin, S., Liu, Y., Yuan, B., Feng, J. Q., & Aswath, P. B. (2015, Dec). The in vivo role of DMP-1 and serum phosphate on bone mineral composition. *Bone*, 81, 602-613. <https://doi.org/10.1016/j.bone.2015.08.018>
- Malaval, L., Wade-Guéye, N. y. M. m., Boudiffa, M., Fei, J., Zirngibl, R., Chen, F., Laroche, N., Roux, J.-P., Burt-Pichat, B., & Duboeuf, F. (2008). Bone sialoprotein plays a functional role in bone formation and osteoclastogenesis. *The Journal of experimental medicine*, 205(5), 1145-1153.

- Marcellini, S., Henriquez, J. P., & Bertin, A. (2012). Control of osteogenesis by the canonical Wnt and BMP pathways in vivo: cooperation and antagonism between the canonical Wnt and BMP pathways as cells differentiate from osteochondroprogenitors to osteoblasts and osteocytes. *Bioessays*, *34*(11), 953-962.
- Mark, M. P., Prince, C. W., Oosawa, T., Gay, S., Bronckers, A. L., & Butler, W. T. (1987). Immunohistochemical demonstration of a 44-KD phosphoprotein in developing rat bones. *Journal of Histochemistry & Cytochemistry*, *35*(7), 707-715.
- Martin, A., David, V., Laurence, J. S., Schwarz, P. M., Lafer, E. M., Hedge, A.-M., & Rowe, P. S. (2008). Degradation of MEPE, DMP1, and release of SIBLING ASARM-peptides (minhibins): ASARM-peptide (s) are directly responsible for defective mineralization in HYP. *Endocrinology*, *149*(4), 1757-1772.
- Martinek, N., Shahab, J., Sodek, J., & Ringuette, M. (2007, Apr). Is SPARC an evolutionarily conserved collagen chaperone? *J Dent Res*, *86*(4), 296-305. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17384023
- Matelot, M., & Noordermeer, D. (2016). Determination of high-resolution 3D chromatin organization using circular chromosome conformation capture (4C-seq). In *Polycomb Group Proteins* (pp. 223-241). Springer.
- Merkel, A., Chen, Y., & George, A. (2019). Endocytic trafficking of DMP1 and GRP78 complex facilitates osteogenic differentiation of human periodontal ligament stem cells. *Frontiers in physiology*, *10*, 1175.
- Meyer, M. B., Benkusky, N. A., & Pike, J. W. (2014). The RUNX2 Cistrome in Osteoblasts Characterization, Down-Regulation Following Differentiation, and Relationship to Gene Expression. *Journal of Biological Chemistry*, *289*(23), 16016-16031.
- Meyer, M. B., Benkusky, N. A., & Pike, J. W. (2015). Selective distal enhancer control of the Mmp13 gene identified through clustered regularly interspaced short palindromic repeat (CRISPR) genomic deletions. *Journal of Biological Chemistry*, *290*(17), 11093-11107.
- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., & Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*, *147*(5), 1132-1145.
- Morrissey, M. A., Jayadev, R., Miley, G. R., Blebea, C. A., Chi, Q., Ihara, S., & Sherwood, D. R. (2016). SPARC promotes cell invasion in vivo by decreasing type IV collagen levels in the basement membrane. *PLoS genetics*, *12*(2), e1005905.
- Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auviel, L., Beaver, J. E., Chowdhary, B. P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S. N., Milan, D., Ostrander, E. A., Pape, G., Parker, H. G., Raudsepp, T., Rogatcheva, M. B., Schook, L. B., Skow, L. C., Welge, M., Womack, J. E., O'Brien S, J., Pevzner, P. A., & Lewin, H. A. (2005, Jul 22). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science*, *309*(5734), 613-617. <https://doi.org/10.1126/science.1111387>
- Nakashima, K., Zhou, X., Kunkel, G., Zhang, Z., Deng, J. M., Behringer, R. R., & De Crombrughe, B. (2002). The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell*, *108*(1), 17-29.
- Niu, L., Shen, W., Shi, Z., He, N., Wan, J., Sun, J., Zhang, Y., Huang, Y., Wang, W., Fang, C., Li, J., Zheng, P., Cheung, E., Chen, Y., Li, L., & Hou, C. (2020). <https://doi.org/10.1101/2020.04.02.021378>
- Noma, K.-i., Allis, C. D., & Grewal, S. I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science*, *293*(5532), 1150-1155. <https://science.sciencemag.org/content/293/5532/1150.long>

Nomura, S., Wills, A. J., Edwards, D. R., Heath, J. K., & Hogan, B. (1988). Developmental expression of 2ar (osteopontin) and SPARC (osteonectin) RNA as revealed by in situ hybridization. *The Journal of cell biology*, 106(2), 441-450.

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., & Duboule, D. (2011, Oct 14). The dynamic architecture of Hox gene clusters. *Science*, 334(6053), 222-225. <https://doi.org/10.1126/science.1207194>

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., & Sedat, J. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), 381-385.

Ohno, S. (1970). *Evolution by gene duplication*.

Otto, F., Thornell, A. P., Crompton, T., Denzel, A., Gilmour, K. C., Rosewell, I. R., Stamp, G. W., Beddington, R. S., Mundlos, S., & Olsen, B. R. (1997). Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell*, 89(5), 765-771.

Pampena, D. A., Robertson, K. A., Litvinova, O., Lajoie, G., Goldberg, H. A., & Hunter, G. K. (2004). Inhibition of hydroxyapatite formation by osteopontin phosphopeptides. *Biochemical Journal*, 378(3), 1083-1087.

Paz, J., Wade, K., Kiyoshima, T., Sodek, J., Tang, J., Tu, Q., Yamauchi, M., & Chen, J. (2005). Tissue- and bone cell-specific expression of bone sialoprotein is directed by a 9.0 kb promoter in transgenic mice. *Matrix biology*, 24(5), 341-352.

Pelletier, N., Champagne, N., Stifani, S., & Yang, X.-J. (2002). MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2. *Oncogene*, 21(17), 2729-2740.

Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell*, 137(7), 1194-1211. [https://www.cell.com/cell/pdf/S0092-8674\(09\)00699-0.pdf](https://www.cell.com/cell/pdf/S0092-8674(09)00699-0.pdf)

Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., Ong, C.-T., Hookway, T. A., Guo, C., & Sun, Y. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6), 1281-1295. [https://www.cell.com/cell/pdf/S0092-8674\(13\)00529-1.pdf](https://www.cell.com/cell/pdf/S0092-8674(13)00529-1.pdf)

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), 279-283.

Rao, S. S., Huang, S.-C., St Hilaire, B. G., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., & Bochkov, I. D. (2017). Cohesin loss eliminates all loop domains. *Cell*, 171(2), 305-320. e324.

Reznikov, N., Hoac, B., Buss, D., Addison, W., Barros, N., & McKee, M. (2020). Biological stenciling of mineralization in the skeleton: Local enzymatic removal of inhibitors in the extracellular matrix. *Bone*, 115447.

Robling, A. G., & Bonewald, L. F. (2020). The osteocyte: new insights. *Annual Review of Physiology*, 82, 485-506.

Sakuma, T., Higashibata, Y., Kawahata, H., Yamada, S., Okabe, M., Kitamura, Y., & Nomura, S. (2003). Difference of osteopontin gene regulation between bone and kidney. *Journal of orthopaedic science*, 8(3), 361-366.

- Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics*, 1.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), 458-472.
- Shlyueva, D., Stampfel, G., & Stark, A. (2014, Apr). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*, 15(4), 272-286. <https://doi.org/10.1038/nrg3682>
- Silvent, J., Sire, J.-Y., & Delgado, S. (2013). The dentin matrix acidic phosphoprotein 1 (DMP1) in the light of mammalian evolution. *Journal of molecular evolution*, 76(1-2), 59-70.
- Sivakumar, A., de las Heras, J. I., & Schirmer, E. C. (2019). Spatial genome organization: from development to disease. *Frontiers in Cell and Developmental Biology*, 7, 18.
- Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., Van Den Heuvel, A., Kockx, C., Palstra, R.-J., Wendt, K. S., Grosveld, F., & Van Ijcken, W. (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nature protocols*, 8(3), 509-524.
- Stein, G. S., Lian, J. B., Van Wijnen, A. J., Stein, J. L., Montecino, M., Javed, A., Zaidi, S. K., Young, D. W., Choi, J.-Y., & Pockwinse, S. M. (2004). Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. *Oncogene*, 23(24), 4315-4329.
- Sumiyama, K., & Tanave, A. (2020). The regulatory landscape of the Dlx gene system in branchial arches: Shared characteristics among Dlx bigene clusters and evolution. *Development, Growth & Differentiation*.
- Tai, P. W., Wu, H., Van Wijnen, A. J., Stein, G. S., Stein, J. L., & Lian, J. B. (2017). Genome-wide DNase hypersensitivity, and occupancy of RUNX2 and CTCF reveal a highly dynamic gene regulome during MC3T3 pre-osteoblast differentiation. *PLoS One*, 12(11), e0188056.
- Tena, J. J., Alonso, M. E., de la Calle-Mustienes, E., Splinter, E., de Laat, W., Manzanares, M., & Gomez-Skarmeta, J. L. (2011). An evolutionarily conserved three-dimensional structure in the vertebrate Irx clusters facilitates enhancer sharing and coregulation. *Nat Commun*, 2, 310. <https://doi.org/10.1038/ncomms1301>
- Thongjuea, S. (2015). r3Cseq: an R package for the discovery of long-range genomic interactions with chromosome conformation capture and next-generation sequencing data.
- Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E., & Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic acids research*, 41(13), e132-e132.
- Touceda-Suárez, M., Kita, E. M., Acemel, R. D., Firbas, P. N., Magri, M. S., Naranjo, S., Tena, J. J., Gómez-Skarmeta, J. L., Maeso, I., & Irimia, M. (2020). Ancient genomic regulatory blocks are a source for regulatory gene deserts in vertebrates after whole genome duplications. *Molecular Biology and Evolution*.
- Vavouri, T., & Lehner, B. (2009). Conserved noncoding elements and the evolution of animal body plans. *Bioessays*, 31(7), 727-735.
- Vavouri, T., McEwen, G. K., Woolfe, A., Gilks, W. R., & Elgar, G. (2006). Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. *Trends in Genetics*, 22(1), 5-10.

Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Korzh, V., Lim, Z. W., Ingham, P. W., Boehm, T., Brenner, S., & Warren, W. C. (2014, May). On the origin of SCPP genes. *Evol Dev*, *16*(3), 125-126. <https://doi.org/10.1111/ede.12072>

Vincent, K., & Durrant, M. C. (2013). A structural and functional model for human bone sialoprotein. *Journal of Molecular Graphics and Modelling*, *39*, 108-117.

Wang, W., Zhong, J., Su, B., Zhou, Y., & Wang, Y.-Q. (2007). Comparison of Pax1/9 locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. *Molecular Biology and Evolution*, *24*(3), 784-791.

Wijchers, P. J., Krijger, P. H., Geeven, G., Zhu, Y., Denker, A., Verstegen, M. J., Valdes-Quezada, C., Vermeulen, C., Janssen, M., & Teunissen, H. (2016). Cause and consequence of tethering a subTAD to different nuclear compartments. *Molecular cell*, *61*(3), 461-473.

Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., Hill, R. E., Bickmore, W. A., & Lettice, L. A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development*, *146*(19), dev179523.

Woltering, J. M., Noordermeer, D., Leleu, M., & Duboule, D. (2014). Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol*, *12*(1), e1001773.

Wong, E. S., Zheng, D., Tan, S. Z., Bower, N. L., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B. M., & Kikuchi, K. (2020). Deep conservation of the enhancer regulatory code in animals. *Science*, *370*(6517).

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., & Kelly, K. (2004). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol*, *3*(1), e7.

Wu, H., Whitfield, T. W., Gordon, J. A., Dobson, J. R., Tai, P. W., van Wijnen, A. J., Stein, J. L., Stein, G. S., & Lian, J. B. (2014). Genomic occupancy of Runx2 with global expression profiling identifies a novel dimension to control of osteoblastogenesis. *Genome biology*, *15*(3), R52.

Ye, L., MacDougall, M., Zhang, S., Xie, Y., Zhang, J., Li, Z., Lu, Y., Mishina, Y., & Feng, J. Q. (2004). Deletion of dentin matrix protein-1 leads to a partial failure of maturation of predentin into dentin, hypomineralization, and expanded cavities of pulp and root canal during postnatal tooth development. *Journal of Biological Chemistry*, *279*(18), 19141-19148.

Zabidi, M. A., & Stark, A. (2016, Dec). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet*, *32*(12), 801-814. <https://doi.org/10.1016/j.tig.2016.10.003>

Zhang, G., & Cohn, M. J. (2008, Aug). Genome duplication and the origin of the vertebrate skeleton. *Curr Opin Genet Dev*, *18*(4), 387-393. <https://doi.org/10.1016/j.gde.2008.07.009>

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in ecology & evolution*, *18*(6), 292-298.

Zhang, J., Tu, Q., & Chen, J. (2009). Applications of transgenics in studies of bone sialoprotein. *Journal of cellular physiology*, *220*(1), 30-34.

ANEXO 1: Partidores

Tabla 1: Secuencias de partidores para diseño de sondas intrónicas

GEN	Partidor	Secuencia	Código LADE
<i>Dmp1</i> -Int	Dm-Int-F5559	tttaacagtaggacagtctgc	P11-I7
	Dm-Int-R6468	agccataatgtaatcagctgg	P11-I9
<i>lbsp</i> -Int	Ib-Int-F8380	aaggaggtggaaagacgatga	P11-J10
	Ib-Int-R9284	tcaaccgtaacaaacaatggg	P12-A2
<i>Scppa2</i> -Exo	Sc-Exo-F251	gagtcatactacaggacctgc	P12-I4
	Sc-Exo-R1179	agctttggctgagttgcatg	P12-I7
<i>Scppa2</i> -Int	Sc2-Int-F4630	ctctacactcccacagaaacc	P12-B7
	Sc2-Int-R5325	attcagaggtgtacacactgg	P12-B9
Sparc-Int	Spr-Int-F2241	tgtagtagtaccgagaaggc	P12-C5
	Spr-Int-R3073	aagggtcgtattgtaatcttc	P12-C7

Tabla 2: Secuencias de partidores para amplificar baits

GEN	Partidor	Secuencia	Código LADE
Ambn	Am-Bait-DpCs-F	ATCCAGAAAATACATGCAGATC	P14C9
	Am-Bait-DpCs-R	CTTCATATGGTATGTCATGTATTTG	P14D1
<i>Dmp1</i>	<i>Dmp1</i> -DpnII-F	CCTTCATCTAGACAATACTAGACATGG	P14A9
	Dm-Tss-F1832	ctttgcaagcatatgagatgc	P11-I3
<i>lbsp</i>	Ib-Bait-DpCs-F	GACTTGCTAATAATGGCAGTG	P14C4
	Ib-Bait-DpCs-R	TCTGTGCTTATTCATTCCTGC	P14C5
<i>Scppa2</i>	Sc2-Tss-F3148	tgcatgtacagtagctgtcata	P12-B2
	Sc2-Tss-R3695	ctgatcccagattccagagc	P12-B6
Sparc	xSparc-DpnII-F	CTGTCCTAACCCCATCCTG	P14B4
	Sp-Bait-DpCs-R	TGCTTCACTTGTACCAGTCC	P14B9

Tabla 3: Secuencias de partidores para PCR inversa

GEN	Partidor	Secuencia	Código LADE
Ambn	xAmbn-Mbol-R	GCAGCTTAATGTTTCAGTGTAATATGAT	914-A6
	xAmbn-Csp6I-F	GAACCACAGACAAAGAATTTTAG	P14-C8
Dmp1	xDmp1-Mbol-R	TCCAGCTTTACAAGGAACCTT	P11-I6
	xDmp1-Csp6I-F	CCTACATGACTTAAGCCCATG	P14-C6
lbsp	xlbsp-Mbol-R	TAATGGGCTATCTAAATTATGTATCTA G	P14-C2
	xlbsp-Csp6I-F	ATAACCTGAATATATATTGCAAATAAG	P14-C3
Scppa2	xScppa2-Mbol-R	GCTCTGGAAATCTGGGATC	P14-B3
	xScppa2-Csp6I-F	AGGCACAATTCTATGACAAGC	P14-C1
Sparc	xSparc-Mbol-R	AGAGTTCTCAGGACAGGATC	P14-B7
	xSparc-Csp6I-F	AAGTTGTTTGTCTGGACTGG	P14-B8

Tabla 4: Secuencias de partidores para PCR inversa con adaptadores

GEN	Partidor	Secuencia
Ambn	P5-xAm-MboR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAGCTTAA TGTTTCAGTGTAATATGAT
	P7i17-xAm-CspF	CAAGCAGAAGACGGCATAACGAGATCTCTACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCGAA CCACAGACAAAGAATTTTAG
Dmp1	P5-xDm-MboR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCAGCTTT ACAAGGAACCTT
	P7i24-xDm-CspF	CAAGCAGAAGACGGCATAACGAGATGCTACC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCCT ACATGACTTAAGCCCATG
lbsp	P5-xBs-MboR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTAATGGGCT ATCTAAATTATGTATCTAG
	P7i26-xBs-CspF	CAAGCAGAAGACGGCATAACGAGATGCTCATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCATA ACCTGAATATATATTGCAAATAAG
Scppa2	P5-xSp2-MboR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTCTGGA AATCTGGGATC
	P7i28-xSp2-CspF	CAAGCAGAAGACGGCATAACGAGATCTTTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCAGG CACAATTCTATGACAAGC
Sparc	P5-xSpr-MboR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGTTCTC AGGACAGGATC
	P7i29-xSpr-CspF	CAAGCAGAAGACGGCATAACGAGATTAGTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCAAG TTGTTTGTCTGGACTGG

Nota: En rojo cada uno de los partidores que amplifican el bait. En azul la etiqueta de identificación *illumina* y en negro la secuencia de los adaptadores P5 y P7.

ANEXO 2: Interacciones con otros cromosomas

Figura 1: Regiones de interacción en *trans* con el promotor *AMBN*

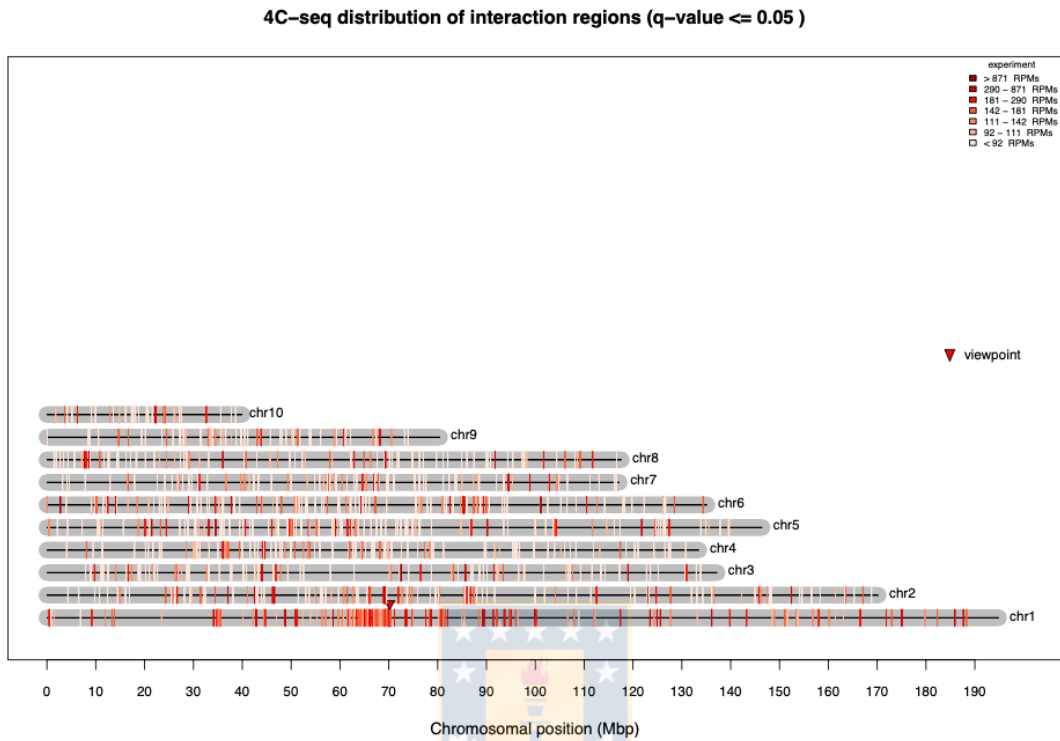


Figura 2: Regiones de interacción en *trans* con el promotor *IBSP*

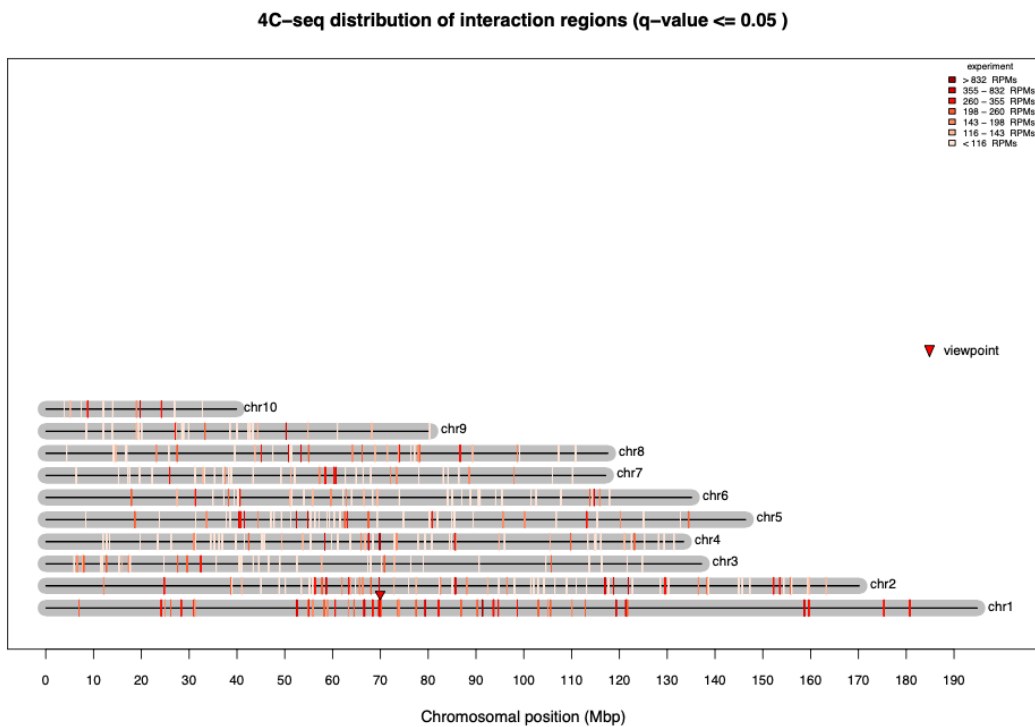


Figura 3: Regiones de interacción en trans con el promotor *DMP1*

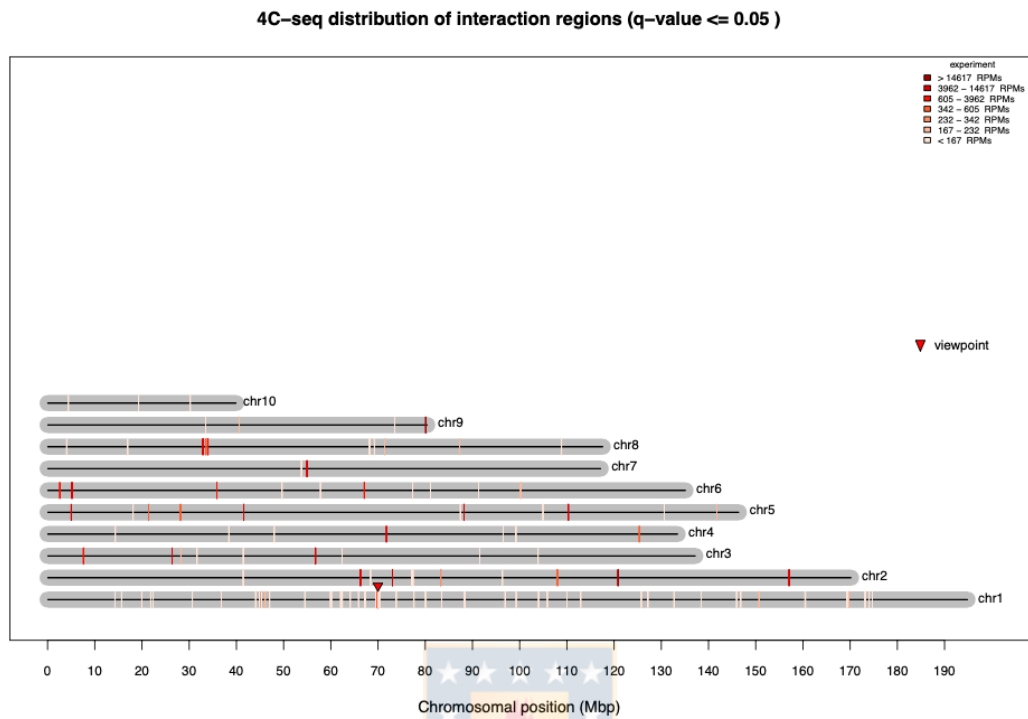


Figura 4: Regiones de interacción en *trans* con el promotor *SCPPA2*

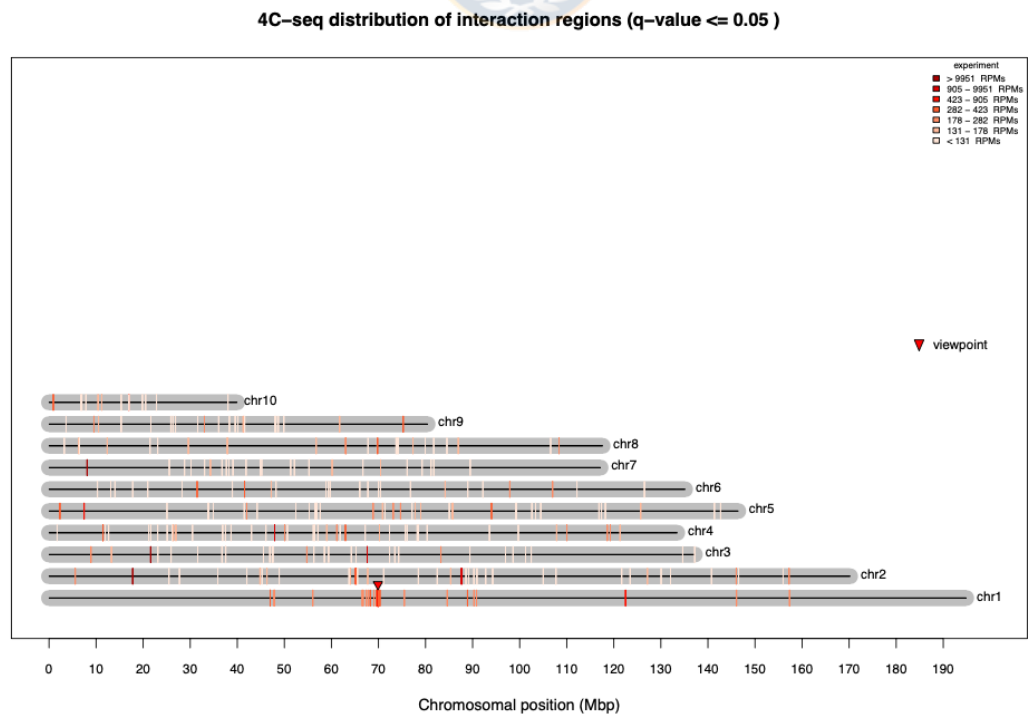
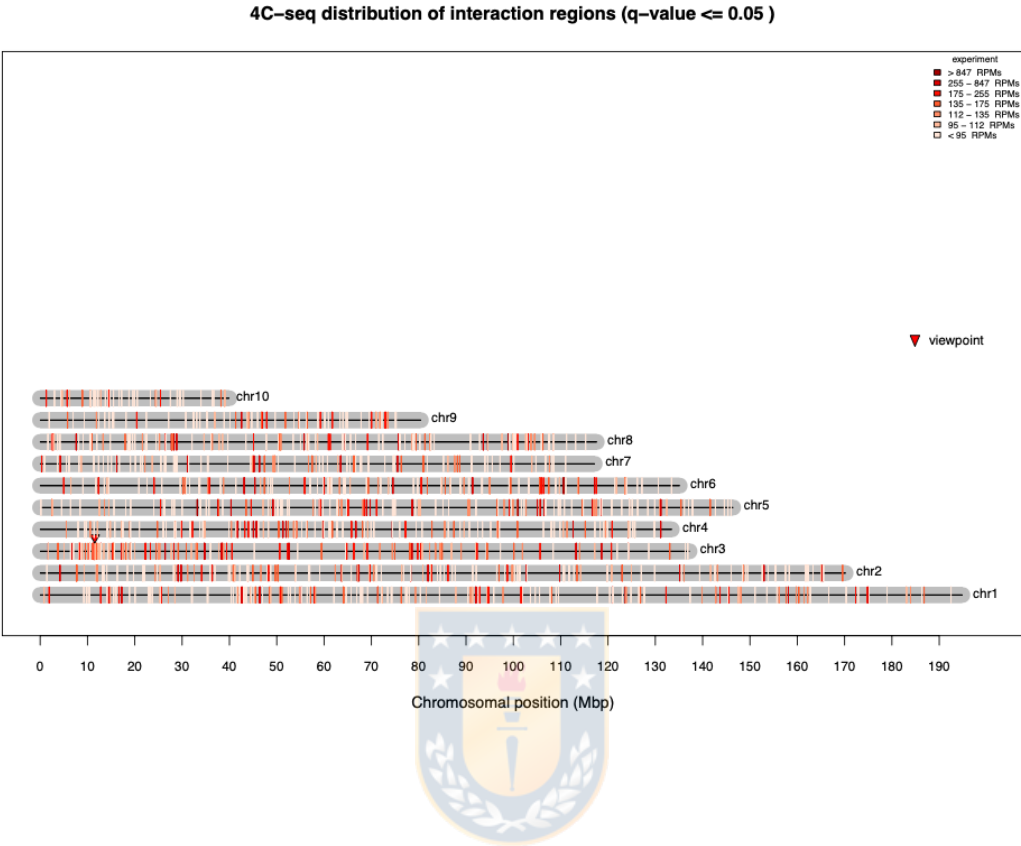


Figura 5: Regiones de interacción en *trans* con el promotor SPARC



ANEXO 3: Locus SCPP y TADs

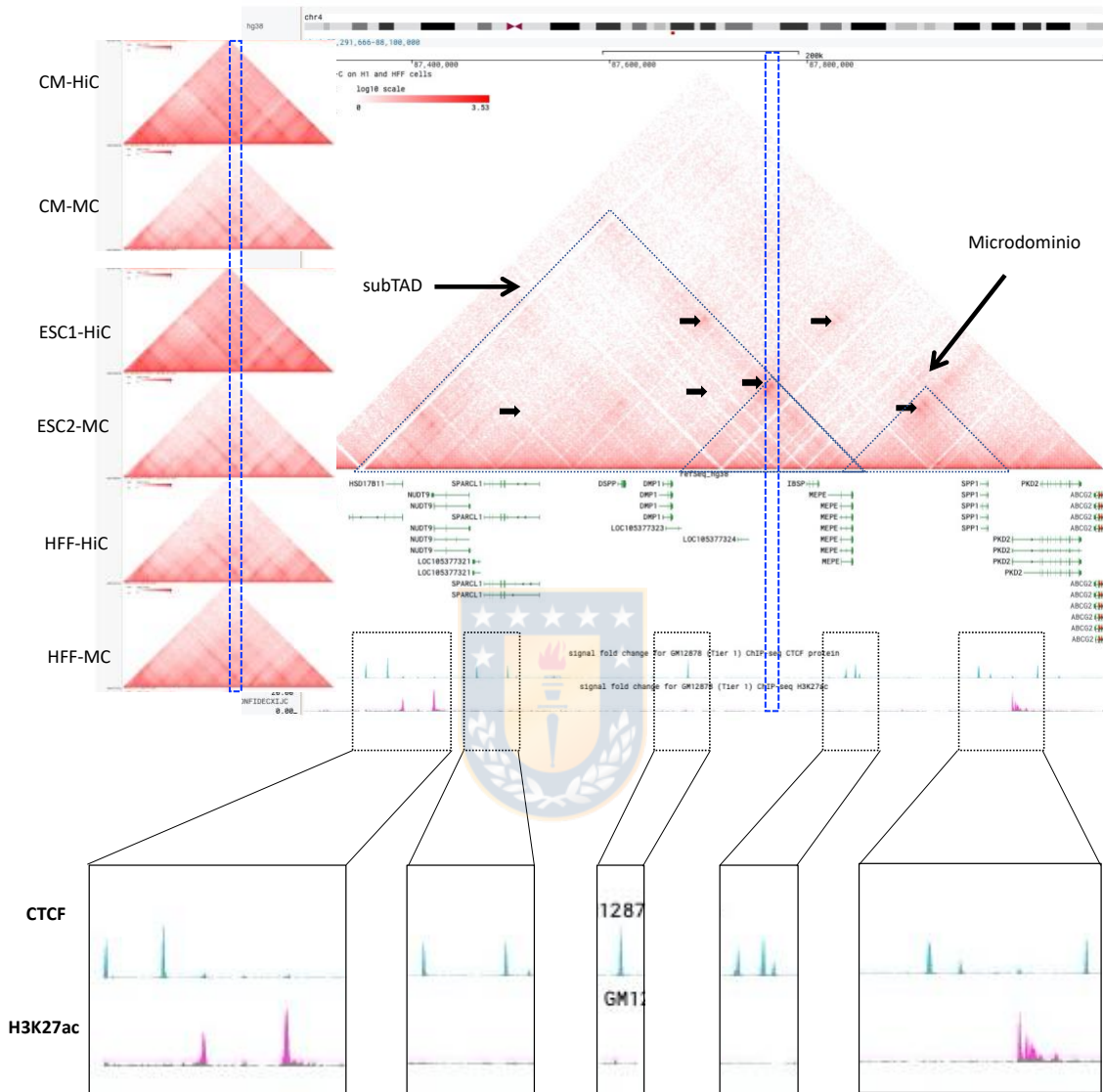


Figura 45: TADs en células humanas asociados al locus SCPP

Los triángulos pequeños representan la estructura del TAD que incluye el locus SCPP en células humanas (CM = Cardiomiocitos; ESC = Células Madre Embrionarias; HFF = Fibroblastos). Se muestra el resultado de 2 técnicas (HiC y Micro-C) para comparar la resolución. Un MicroC ampliado de fibroblastos muestra las frecuencias de interacción. Observe cómo el microdominio (líneas discontinuas azules) se mantiene en la diferenciación celular. Este microdominio incluye a los genes *Dmp1*, *Ibsp* y *MEPE*, los cuáles a su vez están en un subTAD que incluye a *DSPP* y *SPARCL1*. Es interesante observar que *SPP1* se encuentra en otro micro-dominio aunque algunos contactos llegan aún a compartirse (puntos esquinados señalados con flechas). Abajo se muestran los datos de ChIP-seq para CTCF y H3K27ac. Data obtenida en <https://www.4dnucleome.org/>

ANEXO 4: *Sparc* y TADs

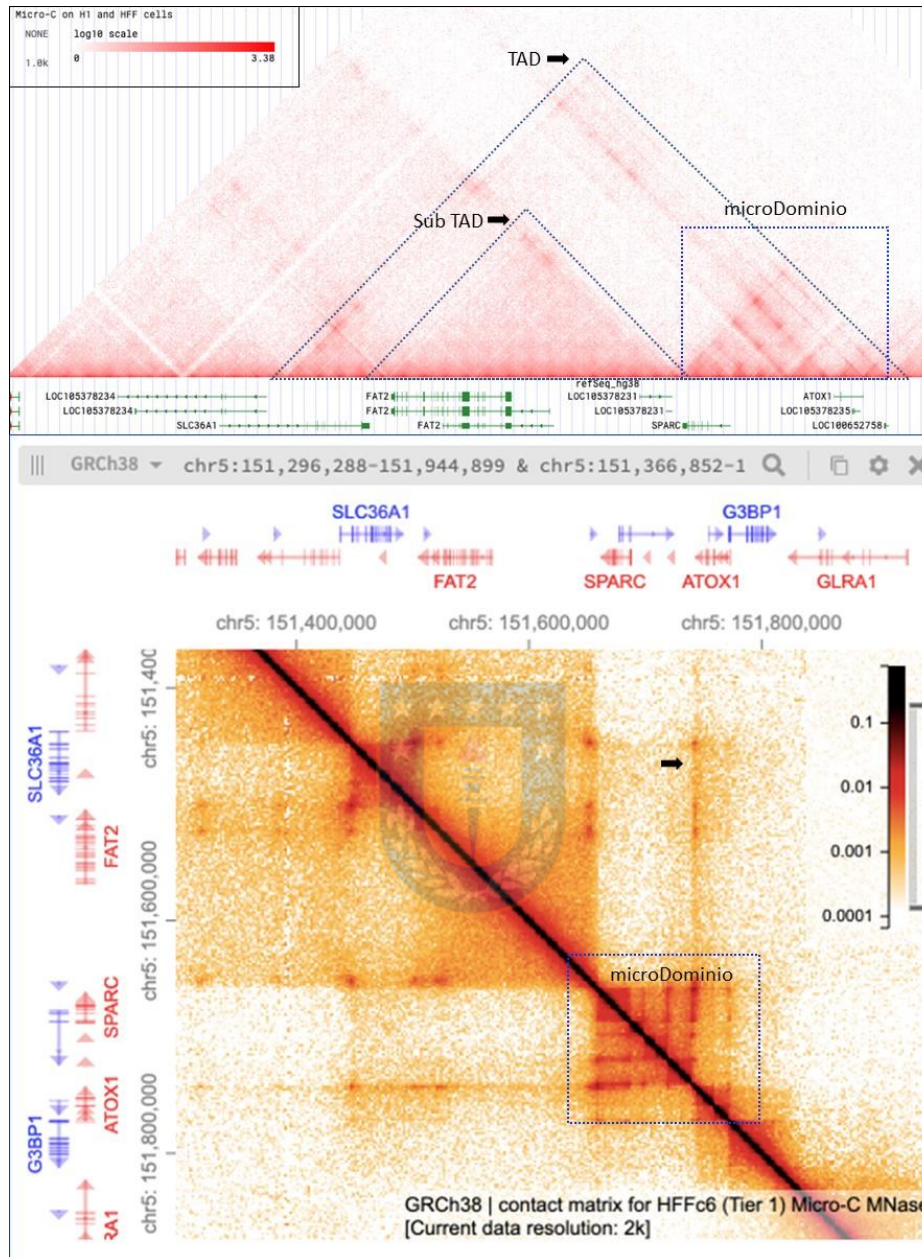


Figura 46: TADs en células humanas asociados al gen SPARC

Sparc se encuentra en un microdominio (fibroblastos). La técnica de MicroC permite observar los contactos específicos en la región río arriba de *Sparc* (zona señalada con un recuadro punteado). La proyección de abajo permite observar un gran punto esquinado que demarca los límites del TAD. Observe como los bystander ATOX1 y FAT2 quedan dentro del TAD pero en diferentes microdominios. Además, un subTAD, señalado con la flecha negra (abajo) separa a ATOX1 pero deja interactuando a FAT2 con SPARC. Esta arquitectura en humanos parece ser similar a lo que sucede en *X. tropicalis*.