



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
INSTITUTO DE FISIOLÓGÍA CELULAR

**Mecanismos moleculares involucrados en la organización
tridimensional del genoma en *Drosophila melanogaster***

TESIS

QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN CIENCIAS

PRESENTA

RODRIGO GACEL ARZATE MEJÍA

DIRECTOR DE TESIS

DR. FÉLIX RECILLAS TARGA

INSTITUTO DE FISIOLÓGÍA CELULAR

COMITÉ TUTOR

DRA. MARTHA VERÓNICA VÁZQUEZ LASLOP

INSTITUTO DE BIOTECNOLOGÍA

DR. JUAN RAFAEL RIESGO ESCOVAR

INSTITUTO DE NEUROBIOLOGÍA

CIUDAD DE MÉXICO, FEBRERO DE 2020



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**Mecanismos moleculares involucrados en la
organización tridimensional del genoma en
*Drosophila melanogaster***

Rodrigo Gacel Arzate Mejía
Ciudad de México
Febrero de 2020



Reconocimientos Académicos

Esta tesis fue realizada bajo la tutoría del Dr. Félix Recillas Targa en el Laboratorio 122 Norte, del Departamento de Genética Molecular del Instituto de Fisiología Celular perteneciente a la Universidad Nacional Autónoma de México como parte de los requisitos para obtener el grado de Doctor en Ciencias a través del Programa de Doctorado en Ciencias Biomédicas.

El comité tutor que evaluó el desarrollo de este proyecto fue integrado por:

Dr. Felix Recillas Targa	Instituto de Fisiología Celular, UNAM
Dra. Martha Vázquez Laslop	Instituto de Biotecnología, UNAM
Dr. Juan Rafael Riesgo Escovar	Instituto de Neurobiología, UNAM

El desarrollo de este trabajo contó con la participación fundamental de la Dra. Mayra Furlan Magaril del Instituto de Fisiología Celular de la UNAM así como con el apoyo del Dr. Victor Corces de la Universidad de Emory.

Se reconoce el apoyo y asistencia técnica de la Bióloga Georgina Guerrero Avendaño en el establecimiento del modelo de *Drosophila melanogaster* en el laboratorio así como su constante apoyo en la realización de experimentos. Así mismo, se reconoce el apoyo fundamental de la laboratorista Gianelli Cortés González.

Se reconoce el apoyo de la Dra. Laura Ongay Larios, Ma. Guadalupe Codiz Huerta y Minerva Mora Cabrera de la Unidad de Biología Molecular del Instituto de Fisiología Celular.

Se reconoce el seguimiento y apoyo a mi situación académica de parte de la Dra. Soledad Funes Argüello y Sara Méndez Ibáñez en la Coordinación de Enseñanza del Instituto de Fisiología Celular.

Se agradece el apoyo constante de la Licenciada Gabriela Valdés Silva del Departamento de Genética Molecular del Instituto de Fisiología Celular.

La realización de este proyecto fue posible gracias al financiamiento proporcionado al Dr. Félix Recillas por el Consejo Nacional de Ciencia y Tecnología (CONACYT) a través del donativo Fronteras de la Ciencia 2015-290, la Fundación Miguel Alemán y por el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (DGAPA-PAPIIT IN203917, IN203620). Durante mis estudios de doctorado gocé de una beca otorgada por CONACTY con número de becario 288814.

Finalmente, se agradece al Jurado para el Examen Doctoral designado por el Programa de Doctorado en Ciencias Biomédicas e intrado por:

Presidente	Dr. David René Romero Camarena	Centro de Ciencias Genómicas
Vocal	Dra. Viviana del Carmen Valadez Graham	Instituto de Biotecnología
Vocal	Dr. Enrique Merino Perez	Instituto de Biotecnología
Vocal	Dr. Felipe Cruz García	Facultad de Química
Secretario	Dr. Félix Recillas Targa	Instituto de Fisiología Celular

La verdadera aristocracia es la de la conciencia

I CHING

El libro de las mutaciones

A mis padres

*con amor incondicional
admiración profunda
y agradecimiento eterno*

“But we as individuals are still hugely blessed. Privileged, and not just privileged to enjoy our planet. More, we are granted to the opportunity to understand why our eyes are open and why they see what they do, in the short time before they close forever”

Unweaving the Rainbow
Richard Dawkins

Índice

Resumen ... 9

Abstract ... 11

Capítulo 1

Una breve historia de la epigenética ... 13

Capítulo 2

Principios de Arquitectura Genómica ... 16

Capítulo 3

Mecanismos de organización 3D del genoma. Vertebrados ... 33

Capítulo 4

Mecanismos de organización 3D del genoma. *D. melanogaster* ... 38

Capítulo 5

Antecedentes directos ... 42

Capítulo 6

Propuesta de investigación ... 45

Capítulo 7

Resultados ... 47

Capítulo 8

Discusión y conclusiones ... 78

Capítulo 9

Perspectivas...86

Bibliografía ... 87

Anexo I.

Métodos ... 98

Tabla 1 y 2 ...104

Anexo II.

Publicaciones generadas durante el Doctorado ... 107

Resumen

En eucariontes el genoma se organiza al interior del núcleo. Al nivel de sub-megabases, los cromosomas se organizan en dominios de alta frecuencia de interacción denominados Dominios Topológicamente Asociados o Dominios Topológicos (Topologically Associating Domains o TADs por sus siglas en inglés) separados uno de otro por fronteras de dominios. La presencia de Dominios Topológicos ha sido descrita en diferentes especies lo cual sugiere que es una característica constante de la organización espacial del genoma al interior del núcleo.

Desde una perspectiva transcripcional, los Dominios Topológicos pueden facilitar el establecimiento de dominios regulatorios al traer en proximidad espacial a elementos regulatorios como promotores y enhancers, mientras que al mismo tiempo, evitan el establecimiento de contactos regulatorios inespecíficos con secuencias fuera de los dominios. La disrupción de Dominios Topológicos a través de la manipulación genética de sus fronteras puede resultar en el establecimiento de contactos regulatorios ectópicos que resultan en cambios en la expresión génica y que pueden tener importantes efectos durante el desarrollo o en patologías como cáncer.

En mamíferos, las fronteras de Dominios Topológicos están ocupadas por Proteínas Arquitectónicas de unión al ADN, como lo son CTCF y YY1 y por el complejo de las Cohesinas. La manipulación genética o epigenética de los motivos de unión a CTCF en las fronteras de Dominios Topológicos o bien la eliminación de CTCF o componentes del complejo de Cohesinas resulta en la pérdida de Dominios Topológicos y alteraciones en la expresión de genes.

En *Drosophila*, las fronteras de Dominios Topológicos coinciden con regiones accesibles a la cromatina que presentan enriquecimiento para marcas post-traduccionales de histonas como H3K27ac y H4K16ac y son sitios de unión para Proteínas Arquitectónicas como CP190, BEAF-32, M1BP, Su(Hw), CTCF, entre otras. Análisis de datos de Hi-C derivados de líneas celulares de *Drosophila* sugieren que los Dominios Topológicos emergen a consecuencia de la actividad transcripcional de secuencias al interior de los dominios, mientras que los sitios genómicos ocupados por Proteínas Arquitectónicas modulan las interacciones físicas entre los dominios adyacentes. La inhibición global de la transcripción en embriones y líneas celulares en *D. melanogaster* resultan en una pérdida significativa de aislamiento entre Dominios Topológicos, sin embargo estos no desaparecen. Por otra parte, la eliminación de la Proteína Arquitectónica BEAF-32 solo resulta en cambios menores en la organización tridimensional del genoma. Por lo tanto, es aún sujeto de intenso debate si los Dominios Topológicos son el resultado de un proceso de agregación de regiones de cromatina que comparten el mismo tipo de estructura de la cromatina o un perfil transcripcional similar o bien surgen como consecuencia de la actividad de elementos genéticos discretos que son reconocidos por Proteínas Arquitectónicas.

El locus de *Notch* en *D. melanogaster* proporcionó la primera evidencia que liga a la organización de la cromatina con el control de la expresión génica. La eliminación de la una secuencia de ~900 pb de la región río arriba del promotor de *Notch* en la mutante *fa(swb)* resulta en la pérdida de la interbanda 3C6-7 y en la fusión de la banda que contiene a *Notch* con la banda río arriba. Este efecto citológico resulta en una reducción en la actividad de la vía de *Notch* que resulta en un fenotipo en el ojo de las moscas. En *Drosophila*, las bandas observadas en cromosomas politénicos corresponden a Dominios Topológicos y las interbandas tienen una buena correspondencia con las fronteras de dichos dominios, lo cual sugiere que el locus de *Notch* es un Dominio Topológico y que su región 5' intergénica corresponde a una frontera de dominio.

En el presente trabajo, investigamos la contribución de las fronteras de dominio a la organización del genoma en TADs y su impacto en la regulación de la expresión génica usando a *Notch* como modelo de estudio en *Drosophila*. Encontramos que la eliminación de las fronteras de dominio da como resultado la fusión de TADs y defectos topológicos de largo alcance que están acompañados por la pérdida de unión de Proteínas Arquitectónicas como CTCF y la ARN Pol II, así como provoca defectos en la transcripción, no tan solo de *Notch*, sino de múltiples genes adyacentes. Nuestros resultados proporcionan evidencia funcional directa sobre la contribución de secuencias genéticas discretas unidas por Proteínas Arquitectónicas y ARN Pol II en la partición de la cromatina en Dominios Topológicos y en la regulación de la expresión génica en *Drosophila melanogaster*.

Abstract

In eukaryotes the genome is organized inside the nucleus. At the level of submegabases, chromosomes are organized into high frequency interaction domains called Topologically Associated Domains (TADs) or Topological Domains separated from each other by domain boundaries. The presence of Topological Domains has been described in different species and therefore is a constant feature of the spatial organization of the genome within the nucleus.

From a transcriptional perspective, Topological Domains can facilitate the establishment of regulatory domains by bringing regulatory elements such as promoters and enhancers in spatial proximity, while at the same time preclude the establishment of nonspecific regulatory contacts with sequences outside the domains. The disruption of topological domains through the genetic transformation of their borders can result in the establishment of ectopic regulatory contacts that affect changes in genetic expression and that can have important effects during development or in pathologies such as cancer.

In mammals, the boundaries of Topological Domains are occupied by Architectural Proteins binding to DNA, such as CTCF and YY1 and by the Cohesin complex. Genetic or epigenetic manipulation of CTCF binding motifs at the borders of Topological Domains or the elimination of CTCF or components of the Cohesin complex results in the loss of Topological Domains and alterations in gene expression.

In *Drosophila*, the boundaries of Topological Domains coincide with chromatin-accessible regions that have enrichment for histone post-translational modifications such as H3K27ac and H4K16ac and are binding sites for Architectural Proteins such as CP190, BEAF-32, M1BP, Su(Hw) , CTCF, among others. Analysis of Hi-C data derived from *Drosophila* cell lines suggests that Topological Domains emerge as a result of the transcriptional activity of sequences within the domains, while genomic sites occupied by Architectural Proteins modulate physical interactions between adjacent domains. Global inhibition of transcription in embryos and cell lines in *D. melanogaster* results in a significant loss of insulation between Topological Domains, however they do not disappear. On the other hand, the removal of the Architectural Protein BEAF-32 only results in minor changes in the three-dimensional organization of the genome. Therefore, it is still subject to intense debate if Topological Domains are the result of an aggregation process of chromatin regions that share the same type of chromatin structure or a similar transcriptional profile or arise as a consequence of the activity of discrete genetic elements that are recognized by Architectural Proteins.

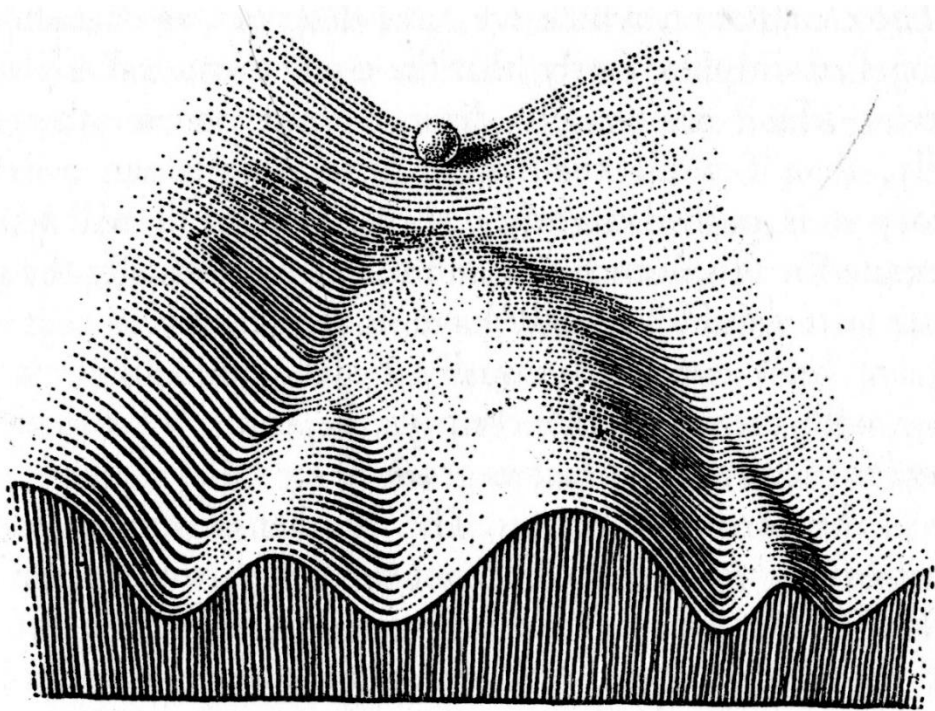
The *Notch* locus in *D. melanogaster* provided the first evidence that links the organization of chromatin with gene expression control. The removal of a ~ 900 bp sequence from the upstream region of the *Notch* promoter in the *fa(swb)* mutant results in the loss of the 3C6-7 interband and in the fusion of the band containing *Notch* with the band upstream. This cytological effect results in a reduction in the activity of the *Notch* pathway that results in a phenotype in the eye of the flies. In *Drosophila*, the bands observed in polytene chromosomes correspond to Topological Domains and the interbands have a good correspondence with the boundaries of these domains, which suggests that the *Notch* locus is a Topological Domain and that its intergenic 5' region corresponds to a boundary of domain.

Here, we investigate the contribution of domain boundaries to genome organization into Topological Domains and their impact on the regulation of gene expression using *Notch* as a study model in *Drosophila*. We found that the elimination of domain boundaries results in the fusion of TADs and long-range topological defects that are accompanied by the loss of binding of Architectural Proteins such as CTCF and RNA Pol II, as well as cause transcription defects, not only at *Notch*, but at multiple adjacent genes. Our results provide direct functional evidence on the contribution of discrete genetic sequences occupied by Architectural Proteins and ARN Pol II in the partition of chromatin in Topological Domains and in the regulation of gene expression in *Drosophila melanogaster*.

Capítulo 1.

Una breve historia de la epigenética

"Epigenetics is founded upon trying to understand unexpected observations"



The epigenetic landscape
Conrad Waddington
1957

El término epigenética -acuñado por el biólogo del desarrollo Conrad Waddington en la década de los 1950s- fue empleado inicialmente para hacer referencia a todos aquellos procesos regulados que, iniciando con el material genético, dan lugar a un organismo. En ese entonces, el término epigenética hacía referencia directa a aquellos procesos moleculares -cuya naturaleza mecánica era desconocida- que durante el desarrollo temprano permitían a una célula adquirir diferentes fenotipos o estados de diferenciación sin la modificación aparente del genotipo. Sin embargo, el concepto de epigenética tiene en realidad su origen en estudios de embriología del siglo XIX, los cuales establecieron las bases de nuestro entendimiento sobre la relación que existe

entre los genes y el desarrollo. En ese entonces existía un fuerte debate sobre la naturaleza y la localización de los componentes que permitían llevar a cabo el plan de desarrollo de un organismo. En particular, un grupo de embriólogos postulaba que el desarrollo embrionario de un organismo era la consecuencia de reacciones químicas entre los componentes solubles del cigoto, a lo cual se conoció como epigénesis.

A finales del siglo XIX e inicios del siglo XX importantes descubrimientos fundaron las bases del campo de la epigenética. El descubrimiento de los cromosomas por Walther Flemming en 1873 y la evidencia de que el programa de desarrollo de los organismos radica en los cromosomas (Wilson y Boveri) proporcionaron las bases conceptuales para pensar en el desarrollo embrionario de un organismo como un proceso cuya ejecución se encontraba fuertemente influido por elementos contenidos al interior del núcleo y fue el trabajo de Thomas Hunt Morgan en 1911 en *Drosophila* que proporcionó evidencia contundente sobre eso al demostrar que los genes podían residir en sitios discretos de los cromosomas¹. Posteriormente, las observaciones de H. J. Muller en 1930 sugirieron que defectos en el desarrollo en *Drosophila*, como por ejemplo la variación del color del ojo, podían ser explicados por mutaciones que no afectaban el contenido de los cromosomas pero si la posición de los elementos que los conformaban, proporcionando la primera evidencia de que se podía alterar el fenotipo de un organismo sin modificación aparente del contenido de cromosoma². El tipo de mutaciones identificadas por Muller se conocen como inversiones y fue hacia la década de los 50s cuando se descubrió que los efectos de variegación en el color de los ojos descritos por Muller surgían en realidad por rearrreglos que situaban el gen *white* al lado de regiones cromosómicas altamente compactas conocidas como heterocromatina, lo cual ahora sabemos influenciaba su estado transcripcional³. Estas observaciones sugirieron que la actividad de los genes no era completamente independiente y que su posición en el genoma podía afectar su función.

Fue hacia la década de 1970 cuando se comenzaron a describir a nivel molecular los primeros mecanismos epigenéticos. En particular, el proceso de inactivación del cromosoma X en el ratón proporcionó uno de los primeros modelos para estudiar mecanismos epigenéticos que no involucraban rearrreglos cromosomales -como los descritos por Muller en 1930- sino que dependían de modificaciones químicas a la molécula del ADN como es la metilación de los dinucleótidos CpG. En los siguientes años se describió la existencia de grupos de genes denominados *Polycomb* y *Trithorax* en *Drosophila* que estaban implicados en modificar de forma permanente el estado de actividad de un gen durante el desarrollo. En ambos casos, las modificaciones se heredaban con cada división celular lo cual sugería un tipo de mecanismo epigenético que afectaba de forma estable el estado transcripcional de los genes.

En esa misma década se dieron avances notables en nuestro entendimiento de la estructura de la cromatina. Por una parte, experimentos llevados a cabo por Cedar y Felsenfeld en 1973⁴ sugerían que solo una fracción del ADN en la cromatina era accesible a la actividad enzimática y Kornberg y Thomas en 1974⁵ describían la existencia del nucleosoma, un octámero de histonas al cual se asocia la molécula de

ADN y que representa la unidad fundamental de la cromatina. Además, la estructura cristalográfica del nucleosoma asociado a la molécula del ADN proporcionó evidencia de la extensión de los extremos amino-terminal de las histonas y sugirió que podrían ser sujetas de modificaciones pos-traduccionales⁶. Hacia 1980, Grunstein y colaboradores proporcionaron evidencia genética en levadura sobre la importancia de las colas de histonas en el control de la expresión génica y el establecimiento de dominios de silenciamiento⁷. Fue en 1996 cuando Allis y colaboradores demostraron que una acetiltransferasa de histonas en *Tetrahymena* era homóloga a un regulador transcripcional de levadura⁸, proporcionando una evidencia contundente de la conexión entre la cromatina, sus modificaciones y el estado transcripcional de los genes y sentando las bases conceptuales de lo que ahora conocemos como epigenética.

El significado del término epigenética propuesto por Waddington hace más de 60 años ha sufrido importantes cambios y esto responde justamente a nuestro avance en la comprensión de los mecanismos moleculares que controlan la expresión génica en eucariontes. Actualmente, el término epigenética es un término sombrilla que hace referencia a una gran variedad de procesos moleculares, que conocemos como mecanismos epigenéticos, y que en conjunto con los factores de transcripción, transducen la información extra celular en cambios en la expresión de genes sin alterar la secuencia de ADN, sino más bien, a través de modificar a la cromatina que es la forma fisiológica de la información genética en eucariontes. De tal forma que, son los mecanismos epigenéticos junto con los factores de transcripción quienes estabilizan los patrones de expresión génica que permiten no tan solo la diferenciación celular durante el desarrollo, sino que proveen a las células de la capacidad de responder y adaptarse ante un medio ambiente cambiante.

Capítulo 2.

Principios de Arquitectura Genómica

Al interior del espacio nuclear, el material genético se asocia con diversas proteínas y moléculas de ARN y en conjunto constituyen a la cromatina. Operacionalmente, la cromatina es el templado sobre el cual ocurren una gran variedad de transacciones nucleares que son fundamentales para la vida celular y la fisiología de los organismos.

Los genomas de eucariontes pueden estar conformados por múltiples cromosomas cada uno formado por una molécula de ADN cuyo tamaño lineal excede el del núcleo que la contiene por órdenes de magnitud. Por ejemplo, el genoma de metazoarios invertebrados como lo es el de la mosca de la fruta, *Drosophila melanogaster*, tiene un tamaño de 180 Mb (megabases)⁹ distribuido en cuatro cromosomas cada uno formado por una molécula de ADN de ~5 cm que se encuentra dentro de un núcleo de un tamaño aproximado de 5 μm , lo cual representa una diferencia de tamaño de 50,000 veces del tamaño del núcleo con respecto a las moléculas de ADN que contiene. Los genomas de animales y plantas pueden alcanzar tamaños excepcionales como el caso del ajolote mexicano (*Ambystoma mexicanum*)¹⁰ de 32 Gb (gigabases) distribuido en 14 cromosomas o de la planta *Paris japonica*¹¹ con un tamaño aproximado de 150 Gb distribuidos en 40 cromosomas. En todos los casos la molécula de ADN en forma de cromatina se encuentra altamente compactada al interior del núcleo, en parte gracias a la asociación con nucleosomas y otras proteínas nucleares. Es en este contexto de compactación que la célula eucarionte debe ejecutar procesos nucleares fundamentales como la replicación y reparación del ADN así como la transcripción. Actualmente sabemos que estos procesos son posibles gracias a la compactación regulada de la cromatina y su organización tridimensional al interior del núcleo.

Observaciones realizadas en la mosca de la fruta en 1930 sugirieron por vez primera una relación entre la arquitectura de los cromosomas y la función del genoma. A lo largo de los últimos 20 años la organización tridimensional de la cromatina al interior del núcleo ha emergido como un aspecto fundamental para el control de diversos procesos nucleares como la regulación transcripcional, la replicación y reparación del ADN, entre otros. Sin embargo, nuestra comprensión sobre los mecanismos implicados en la organización del genoma al interior del núcleo así como su impacto en la regulación de los genes es aún limitada. En esta sección presento un breve resumen del conocimiento actual que tenemos sobre la organización tridimensional del genoma. En particular, profundizaré sobre las características de los Dominios Topológicos.

Evidencia inicial sobre el papel de la arquitectura de los cromosomas en la función del genoma

La primera evidencia experimental que sugirió una función reguladora de la organización de los cromosomas fue la descripción del efecto de variegación por posición (PEV, por sus siglas en inglés) descrito originalmente por Herman Muller en 1930. Muller describió la existencia de efectos en la coloración de los ojos en *Drosophila melanogaster* y que afectaban la función pero no la secuencia del gen *white*². Descubrimientos posteriores basados principalmente en observaciones citológicas y datos genéticos, demostrarían que esto correlacionaba con la yuxtaposición de *white* con regiones de heterocromatina pericentromérica a consecuencia de rearrreglos cromosomales como inversiones³ (Figura 1). Estas observaciones sugerían que, se podía modificar la actividad de un gen tan solo cambiando su posición, ya sea con respecto al mismo cromosoma (inversiones) o al posicionarlo en proximidad lineal con otro cromosoma (translocaciones). Además, se estableció por primera vez la correlación entre “actividad” de un gen y su cercanía a una región de cromatina que se pensaba era inerte (heterocromatina), lo cual significaba un importante avance conceptual en nuestro entendimiento sobre los procesos que regulan la función del genoma.

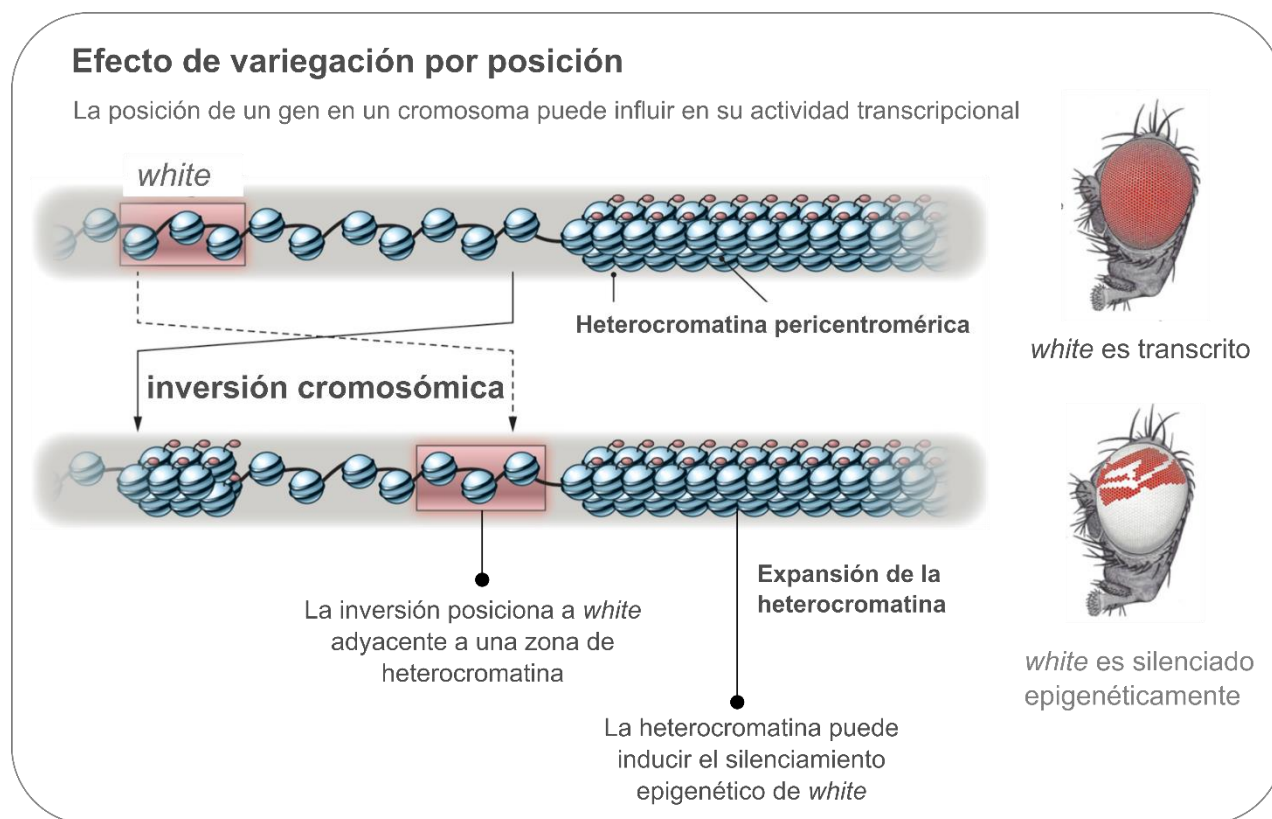


Figura 1. Efecto de variegación por posición. Figura modificada de Brummelkamp and van Steensel 2015 (<https://science.sciencemag.org/content/348/6242/1433/F1>).

Hacia 1980 se popularizó el uso de una técnica conocida como transformación por elementos P la cual hace uso de transposones que se integran de forma semi-azarosa al genoma y fueron ampliamente utilizados para estudiar la función de genes o bien para movilizar a genes reporteros¹². La integración de genes reporteros mediante esta técnica resultaba frecuentemente en diferencias en la actividad del reportero y esto correlacionaba con sus diferentes sitios de integración¹³, lo cual sugería que las secuencias adyacentes al sitio genómico de integración interferían con la regulación del gen reportero. Por lo tanto los hallazgos asociados a la caracterización del efecto de variegación por posición y los obtenidos por las integraciones de elementos P sugerían dos fenómenos cruciales en la regulación transcripcional en eucariontes: por una parte la actividad de los genes podía ser afectada por las secuencias genéticas adyacentes o por su estado cromatínico. Esta observación tenía además una implicación fundamental, y es que como lo expresan Schwarts y Cavalli¹⁴ deberían existir mecanismos regulatorios que protegieran a los genes de estimulación no específica por su entorno cromatínico. Es precisamente con estas observaciones que surge en el campo de la cromatina la idea de la existencia de mecanismos de aislamiento regulatorio o "insulation", una idea que demostraría ser fundamental para nuestra comprensión de los mecanismos de regulación de la expresión génica en metazoarios.

Los efectos de variegación por posición se explican por lo que conocemos como efectos de corta distancia, es decir, la actividad de un gen puede ser influenciada por secuencias que se encuentran adyacentes a sus elementos regulatorios. Sin embargo, la expresión de los genes puede ser también afectada por secuencias que se encuentran a largas distancias o bien incluso en diferentes cromosomas. Fue precisamente en *Drosophila* en la década de 1950 cuando Lewis reportó lo que conocemos como el fenómeno de transvección¹⁵. En *Drosophila* los cromosomas somáticos homólogos se encuentran asociados uno con el otro, de tal forma, que en casos de transvección, secuencias regulatorias de un cromosoma homólogo pueden activar o reprimir la transcripción de genes en el otro cromosoma homólogo. Además, las regiones pericentroméricas de heterocromatina que amplias regiones del genoma (~10Mb) pueden establecer interacciones en *trans* y formar una estructura conocida como centrosoma, la cual puede inducir el silenciamiento de genes que son reposicionados hacia dicha estructura.

Finalmente, el descubrimiento y estudio de los cromosomas politénicos en *Drosophila* proporcionó evidencia fundamental sobre la partición de los cromosomas en estructuras discretas, conocidas como bandas e interbandas, que a nivel citológico sugerían diferencias en su nivel de compactación. El mapeo de genes a bandas de cromosomas politénicos y de elementos de regulación como promotores a sus interbandas sugirió fuertemente una relación funcional entre la organización de los cromosomas y el control de la expresión génica.

Así, fueron estudios pioneros en la mosca de la fruta los que proporcionaron las primeras evidencias sólidas que sugerían que la organización de los cromosomas influía en la actividad de los genes.

Métodos para el estudio de la arquitectura de cromosomas

Los primeros métodos empleados para estudiar la organización de la cromatina al interior del núcleo hicieron uso de técnicas de tinción de cromosomas y su análisis por microscopía óptica. Estos estudios revelaron la organización del genoma en regiones de cromatina accesible (eucromatina) o compacta (heterocromatina). Preparaciones de microscopía electrónica confirmaron la existencia de al menos dos tipos de cromatina que diferían en su compactación. La eucromatina se podía observar con frecuencia hacia el centro del núcleo mientras que la heterocromatina se localizaba con preferencia en la periferia nuclear. La eucromatina se asociaba a regiones de cromosomas permisivas a la transcripción, mientras que la heterocromatina era descrita como un tipo de cromatina inerte, sin actividad transcripcional y conformada por elementos repetidos. Estas observaciones permitieron correlacionar por primera vez el tipo de cromatina (eucromatina o heterocromatina) con su ubicación en el núcleo en dos dimensiones (centro o periferia) y su estado transcripcional (activo o inactivo). Además, fue evidente que la disposición espacial de ambos tipos de la cromatina difería entre tipos celulares, lo cual también sugería que las diferencias transcripcionales entre tipos celulares podría asociarse con cambios en la accesibilidad de la cromatina y su disposición en el espacio nuclear. Hacia el año 2000, el desarrollo de técnicas citológicas para el marcaje fluorescente de cromosomas como la Hibridación fluorescente *in situ* (FISH por sus siglas en inglés) permitió estudiar por primera vez el posicionamiento de cromosomas enteros o regiones genómicas específicas en el espacio nuclear. Los datos de FISH demostraron por primera vez que los cromosomas en el interior del núcleo se ocupan regiones discretas, conocidas como territorios cromosomales, las cuales varían dependiendo del tipo celular. Estos hallazgos demostraban por primera vez que los cromosomas al interior del núcleo no se encuentran en un estado desordenado e interactuando indiscriminadamente, lo cual era difícil de determinar por métodos de microscopía electrónica, y sugerían la existencia de mecanismos moleculares que influían en la localización de los cromosomas. Además, el descubrimiento de territorios cromosomales permitió comprender el origen de ciertos rearrreglos cromosómicos como las translocaciones.

Nuestro entendimiento sobre la organización tridimensional de la cromatina al interior del núcleo avanzó rápidamente en los últimos 20 años gracias al desarrollo de la técnica de Captura Conformacional de Cromosomas (3C) (Figura 2). El 3C fue desarrollado hacia el año 2002 por Job Dekker y es una técnica cuyo objetivo es identificar si dos segmentos de la cromatina se encuentran en proximidad espacial, es decir, si establecen un contacto de larga distancia o asa de cromatina¹⁶. Esta técnica

parte de dos importantes consideraciones. La primera es que los contactos de entre cualesquiera dos segmentos de la cromatina están mediados por proteínas y la segunda que la identificación de estos contactos puede ser posible gracias a una reacción de ligación intra-molecular. En este método, las células son fijadas con formaldehído para así estabilizar los contactos ADN-proteína. Posteriormente se genera una reacción de restricción *in-situ* de genoma completo empleando enzimas de restricción que cuyo tamaño de sitio de reconocimiento varía de 6 a 4 nucleótidos. La elección de la enzima de restricción será determinante para el nivel de resolución de contactos que podamos identificar, de tal forma que una enzima que reconoce motivos de 4 nucleótidos dará el máximo nivel de resolución y podremos identificar contactos que se encuentren a distancias menores (<10 kb). La reacción de digestión tiene como objetivo fragmentar la cromatina fijada y generar extremos cohesivos que posteriormente serán sustratos para una reacción de ligación intra-molecular. La adición de ligasa en concentraciones subóptimas favorece la ligación de segmentos en proximidad espacial, de tal forma que, una vez concluida la ligación, el ADN se purifica y es empleado para su análisis mediante PCR. Un experimento de 3C generará millones de fragmentos de ligación. Una gran proporción serán eventos de re-circularización, pero también muchos más serán eventos de ligación entre fragmentos distantes creando moléculas híbridas las cuales pueden ser identificadas mediante PCR con oligonucleótidos localizadas en cada uno de los fragmentos a analizar.

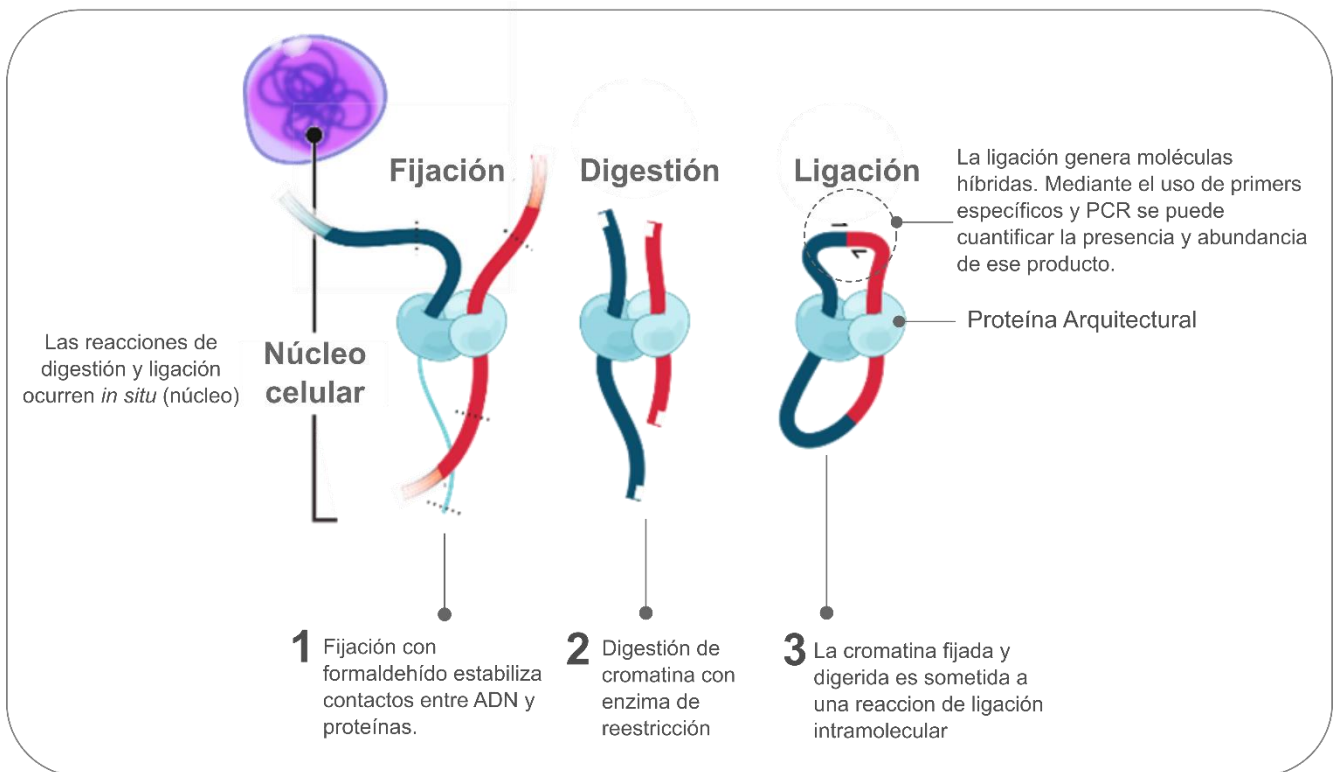


Figura 2. Captura Conformacional de Cromosomas (3C).

Fue la introducción de esta técnica lo que potenció y generalizó el estudio de los contactos de larga distancia que anteriormente estaba limitada al uso de FISH (*Fluorescence in situ hybridization*) una técnica cuya resolución en ese entonces limitaba el estudio de la dinámica nuclear a un par de regiones a la vez y que evaluaba segmentos de la cromatina de gran tamaño (>50 kb) además de otras complejidades técnicas.

En los años siguientes diversas adaptaciones del 3C asociadas a secuenciación masiva, como el 4C, 5C, ChIA-PET o Hi-C revolucionaron nuestro entendimiento sobre la organización 3D del genoma al interior del núcleo¹⁷. El 4C permite identificar todos los elementos que establecen contactos de larga distancia con una región genómica de interés mientras que el 5C permite caracterizar todo el perfil de interacción de una zona del genoma. El ChIA-PET incorpora un paso de inmunoprecipitación después del 3C lo cual permite recuperar aquellas interacciones mediadas por una proteína de interés. Por su parte el Hi-C permite conocer el perfil de interacción de todas las regiones del genoma en un solo experimento. A diferencia de los otros métodos, el Hi-C es un método no sesgado y exhaustivo para caracterizar la organización tridimensional del genoma, ya que no requiere seleccionar regiones particulares para estudiar (3C.4C), tampoco restringe el análisis a una zona del genoma (5C) y no requiere conocer a priori una proteína que pudiera ser responsable de mediar interacciones cromatínicas (ChIA-PET). Es precisamente por estas peculiaridades, que de entre todas ellas, la que tuvo el mayor impacto en el campo fue el Hi-C.

Hacia 2009, Job Dekker y Erez Lieberman Aiden introdujeron la técnica de Hi-C (Figura 3)¹⁸. A diferencia de un 3C convencional, en Hi-C después de la restricción del genoma se introduce un paso de reparación de los extremos cohesivos haciendo uso de un nucleótido biotinilado. Esto genera extremos no cohesivos pero que tienen una molécula de biotina. Posteriormente a la ligación, el material de 3C es sometido a un *pull-down* con perlas de estreptavidina y numerosos lavados que permiten recuperar solo aquellas moléculas que fueron restringidas, reparadas, que incorporaron el nucleótido con biotina y que fueron ligadas, es decir productos de 3C. El material recuperado entonces es usado para secuenciación. La introducción del marcado de moléculas de 3C con biotina y su recuperación con estreptavidina permitió reducir la fracción de ADN no ligado que generalmente acompaña a un experimento de 3C convencional y que por lo tanto resulta en lecturas de secuenciación no útiles y eleva los costos del experimento.

El uso de Hi-C permitía entonces conocer la información espacial de cualesquiera dos fragmentos del genoma en el tipo celular estudiado. Es fundamental señalar que las técnicas basadas en 3C proporcionan información probabilística sobre la cercanía espacial de cualesquiera dos segmentos del genoma. Es decir, dos segmentos de ADN no permanecen en proximidad espacial de forma estática en una célula. De tal forma que un contacto de 3C o Hi-C de alta frecuencia de interacción representa un contacto que es identificado en una gran proporción de las células empleadas en el

experimento. En este sentido, técnicas basadas en microscopía como FISH permiten complementar los resultados de un experimento basado en 3C al proporcionar información sobre la distancia relativa entre dos fragmentos y la distribución célula a célula de las distancias entre dichos fragmentos.

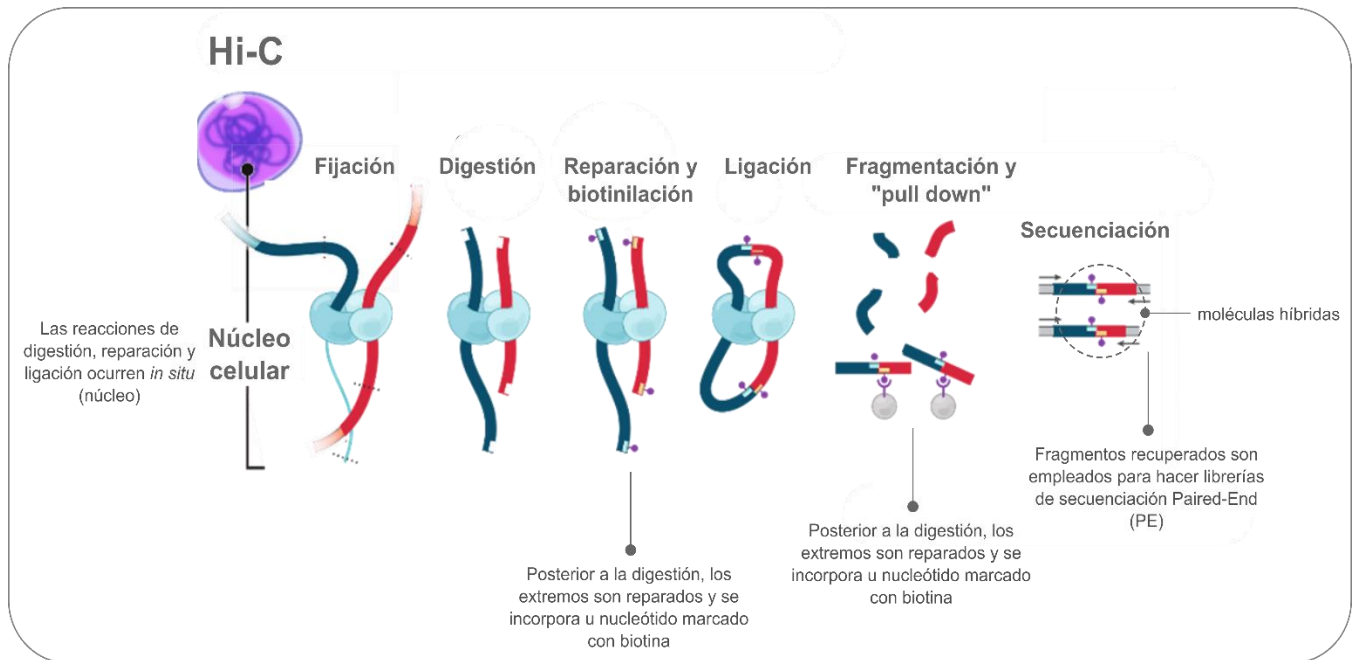


Figura 3. Experimento de Hi-C

A diferencia de un experimento de 3C, los experimentos de Hi-C requieren métodos sofisticados de análisis computacional y de visualización específicos ya que contienen información de la organización tridimensional de todo el genoma. Para la interpretación visual y cuantitativa de experimentos de Hi-C se introdujeron matrices de contacto.

Por ejemplo, si deseamos visualizar la información de contactos para el cromosoma 1 del genoma humano, podemos crear una matriz simétrica en donde cada celda es de un tamaño constante, por ejemplo 100 kb, y representa lo que conocemos como la resolución de la matriz (Figura 4). La primera fila representa segmentos de 100 kb del cromosoma 1 y la primera columna también. A cada celda al interior de la matriz asociamos un valor que representa el número de lecturas de secuenciación asociadas a ese fragmento híbrido. Dada la naturaleza simétrica de la matriz, la diagonal representa los contactos de cada fragmento consigo mismo es decir, representa eventos de re-circularización de los fragmentos digeridos y por lo general tiene el máximo número de fragmentos en los experimentos de Hi-C, esta información no es de utilidad para análisis posteriores. En este ejemplo la celda que corresponde a los fragmentos híbridos contenidos entre los segmentos 101Mb y 102 Mb tiene un valor de

350, es decir, 350 lecturas de secuenciación diferentes asociadas a este fragmento híbridos. Así, cada par de celdas recibe un valor ($n \geq 0$, $n = \text{número de lecturas}$). La matriz de contactos puede entonces ser empleada por programas de visualización para crear mapas de calor que es lo que conocemos como mapas de Hi-C.

Los datos de experimentos de Hi-C revelaron una compleja organización de los cromosomas hasta el momento desconocida¹⁸⁻²⁰. En particular, los mapas de Hi-C mostraron la existencia de regiones del genoma que presentaban una elevada frecuencia de interacción al interior de los cromosomas o bien entre cromosomas, es decir, la cromatina de eucariotes en interfase estaba organizada en dominios de contacto (Figura 4 y 5).

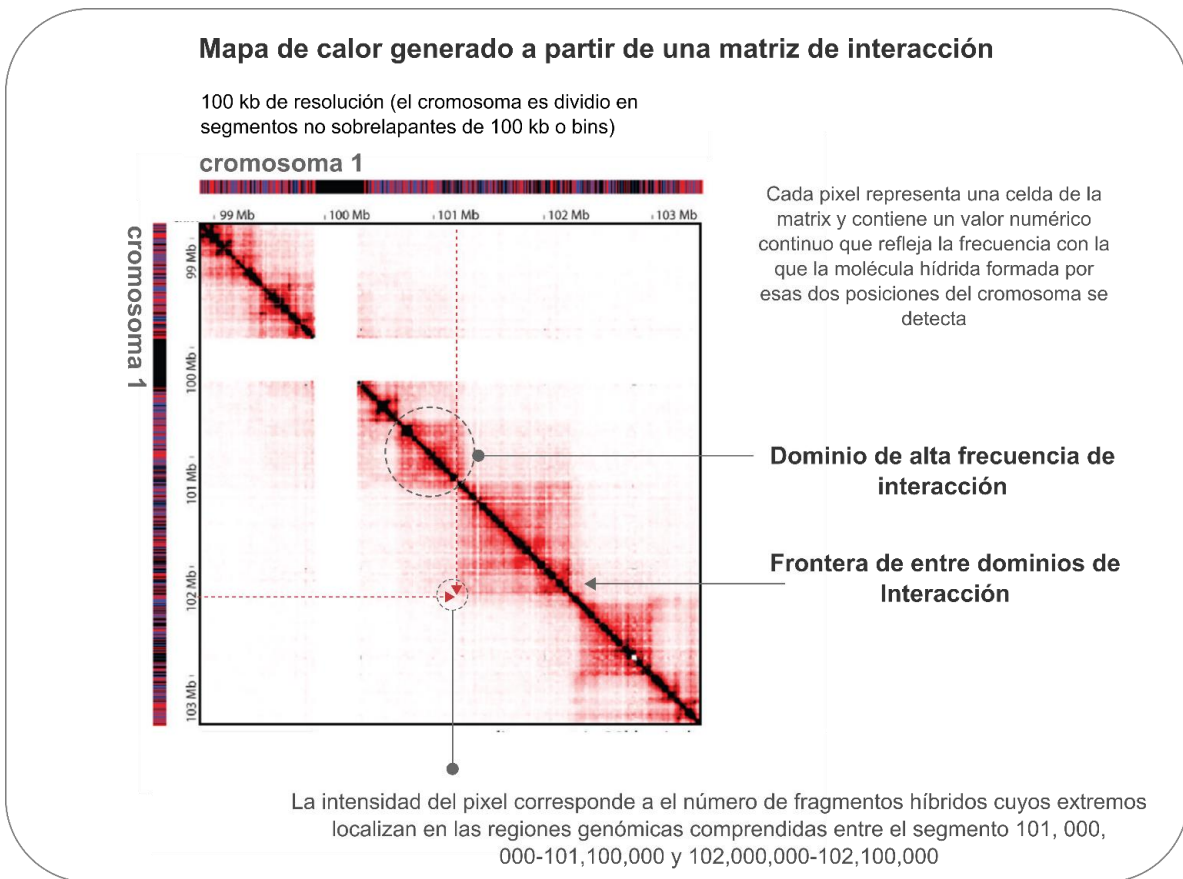
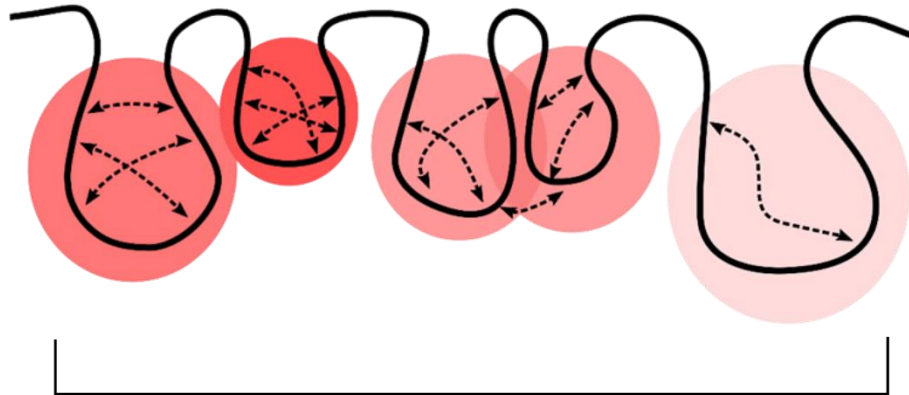


Figura 4. Mapa de calor generado a partir de una matriz de interacción.

Puesto que la matriz de interacción es simétrica, para términos de visualización los datos se pueden mostrar presentando solo la información de un lado de la diagonal



Representación esquemática de los dominios visibles en una matriz de Hi-C

Figura 5. Los mapas de Hi-c revelan la existencia de dominios de elevada frecuencia de interacción.

La organización tridimensional de la cromatina

La cromatina al interior del núcleo se organiza a distintos niveles. Desde la agregación de nucleosomas hasta el posicionamiento de los cromosomas en el espacio nuclear, cada nivel de organización está sujeto de regulación y puede afectar diversos procesos nucleares. A continuación describiré cada uno de los niveles de organización cromosomal en metazoarios y resumiré nuestra comprensión actual de cada uno.

Interacciones entre nucleosomas

El nivel más básico de organización está representado por las interacciones físicas entre nucleosomas conocidas como 'clutches' y que difieren entre tipos celulares, lo cual probablemente refleja diferencias en las modificaciones post-traduccionales de los nucleosomas entre tipos celulares (Figura 6)²¹. Contrario a lo esperado, la fibra de cromatina es flexible y esto se refleja en los arreglos heterogéneos de grupos de nucleosomas que son evidentes por microscopía de super-resolución.

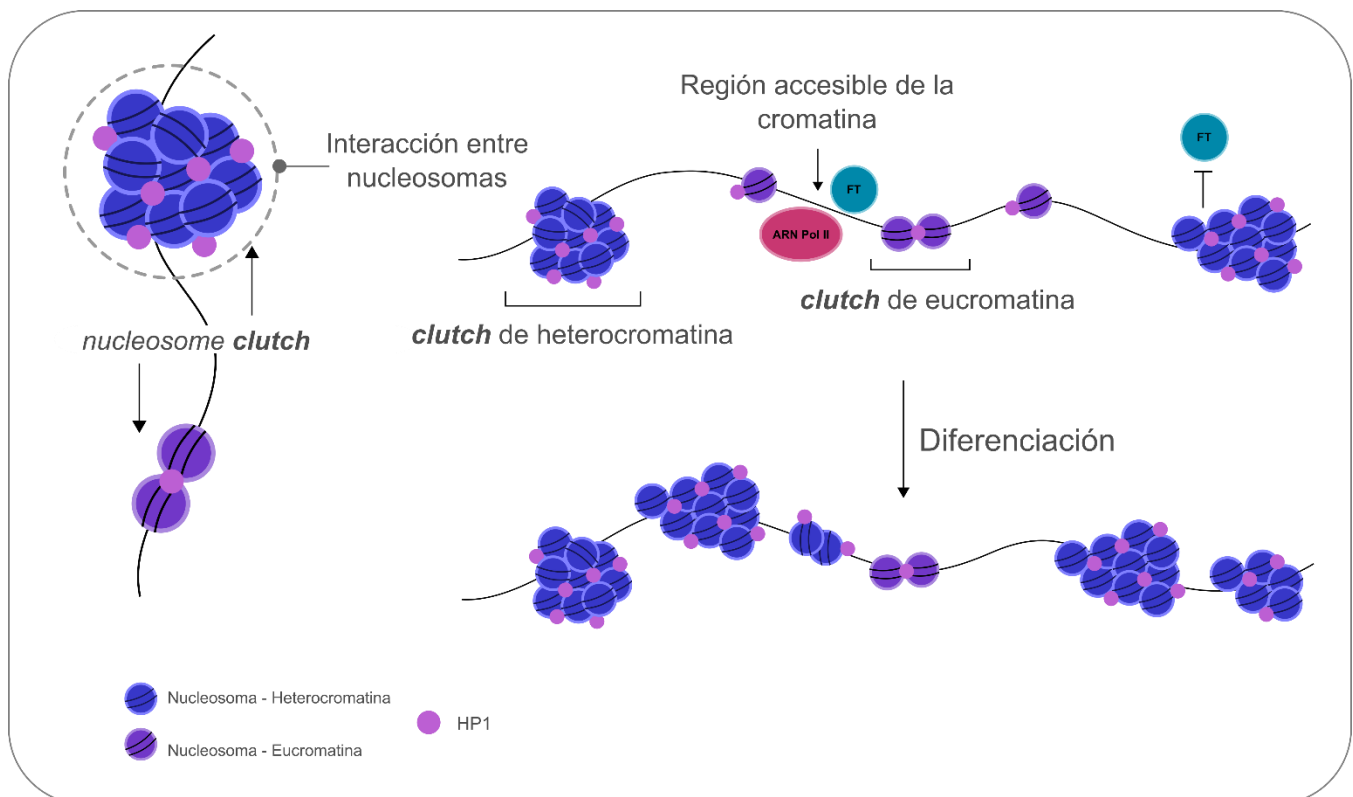


Figura 6. Agregados de nucleosomas.

Asas de cromatina

La fibra de cromatina puede plegarse y generar asas de cromatina^{22,23}. Las asas de cromatina son una característica de los genomas de eucariontes, particularmente de los metazoarios en donde los elementos de regulación en cis pueden encontrarse a grandes distancias de sus promotores blancos en la secuencia lineal de los cromosomas. Por ejemplo, las asas de cromatina mejor caracterizadas son aquellas que se establecen entre elementos enhancer y sus promotores blancos (Figura 7). En este contexto, las asas de cromatina se piensa que proporcionan un nivel de especificidad adicional a la actividad de los elementos enhancer al establecer especificidad de estimulación.

Las asas de cromatina también se pueden dar entre elementos regulatorios del mismo tipo, como por ejemplo, entre enhancers o promotores, lo cual se piensa crea centros de agregación de ARN Pol II y podría favorecer la transcripción, o bien entre secuencias con actividad aisladora que crean un tipo de asa de cromatina conocida como estructural o arquitectónica y se postula, favorece interacciones regulatorias entre secuencias al interior de la asa²². Las asas de cromatina también se pueden establecer entre genes, por ejemplo, aquellos que son co-regulados transcripcionalmente o bien aquellos que son reprimidos epigenéticamente por miembros del grupo Polycomb²⁴. Además, en levadura y en *Drosophila*, los genes activos transcripcionalmente establecen contactos entre el promotor y el final del gen formando un tipo de asa de cromatina conocida como asa de gen o *gene loop* que se ha propuesto refuerza la direccionalidad de la síntesis de ARN por la ARN polimerasa²⁵.

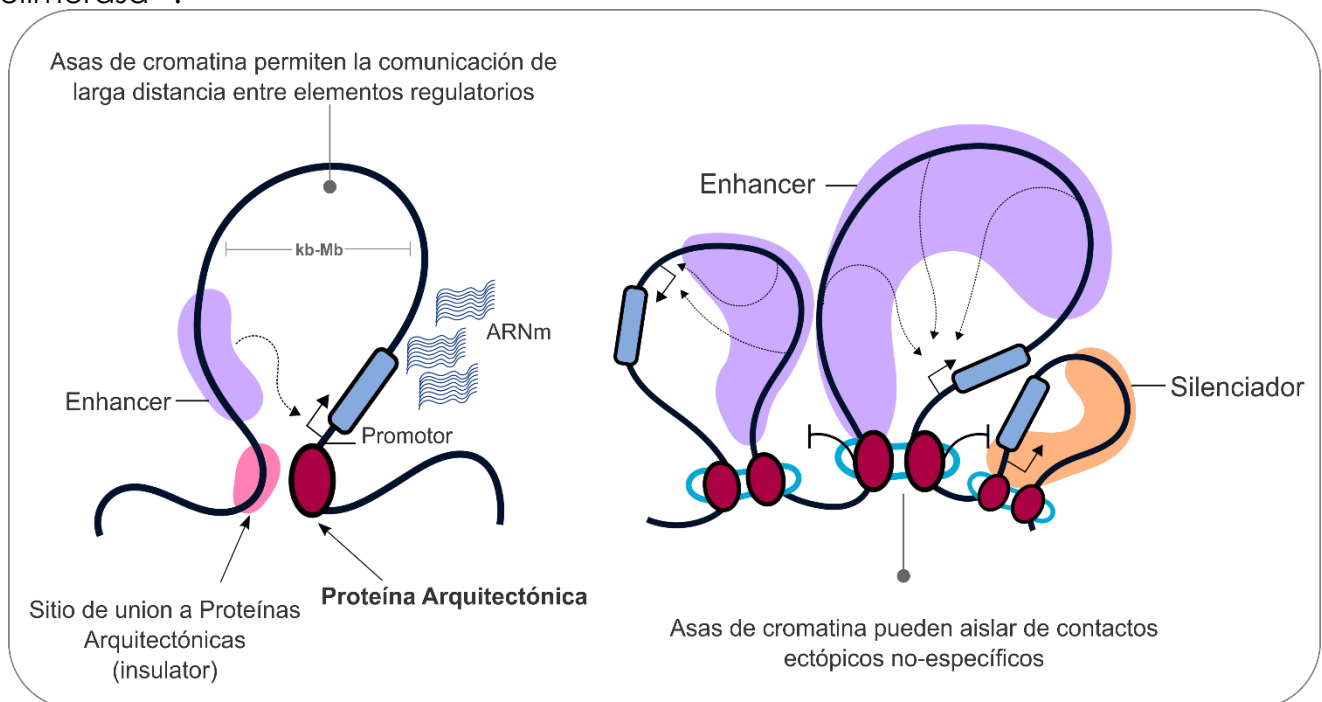


Figura 7. Asas de cromatina como mediadoras de contactos de larga distancia.

A nivel de Hi-C, las asas de cromatina son visualmente identificables como “puntos” en los mapas de calor (Figura 8) y corresponden a regiones de muy alta frecuencia de interacción²⁶. Conceptualmente representan las bases de una asa de cromatina y es que son esos segmentos de la asa los que están en mayor cercanía espacial. En las bases de las asas se encuentran frecuentemente Proteínas Arquitectónicas entre ellas CTCF (vea sección: Mecanismos de organización 3D del genoma: Vertebrados).

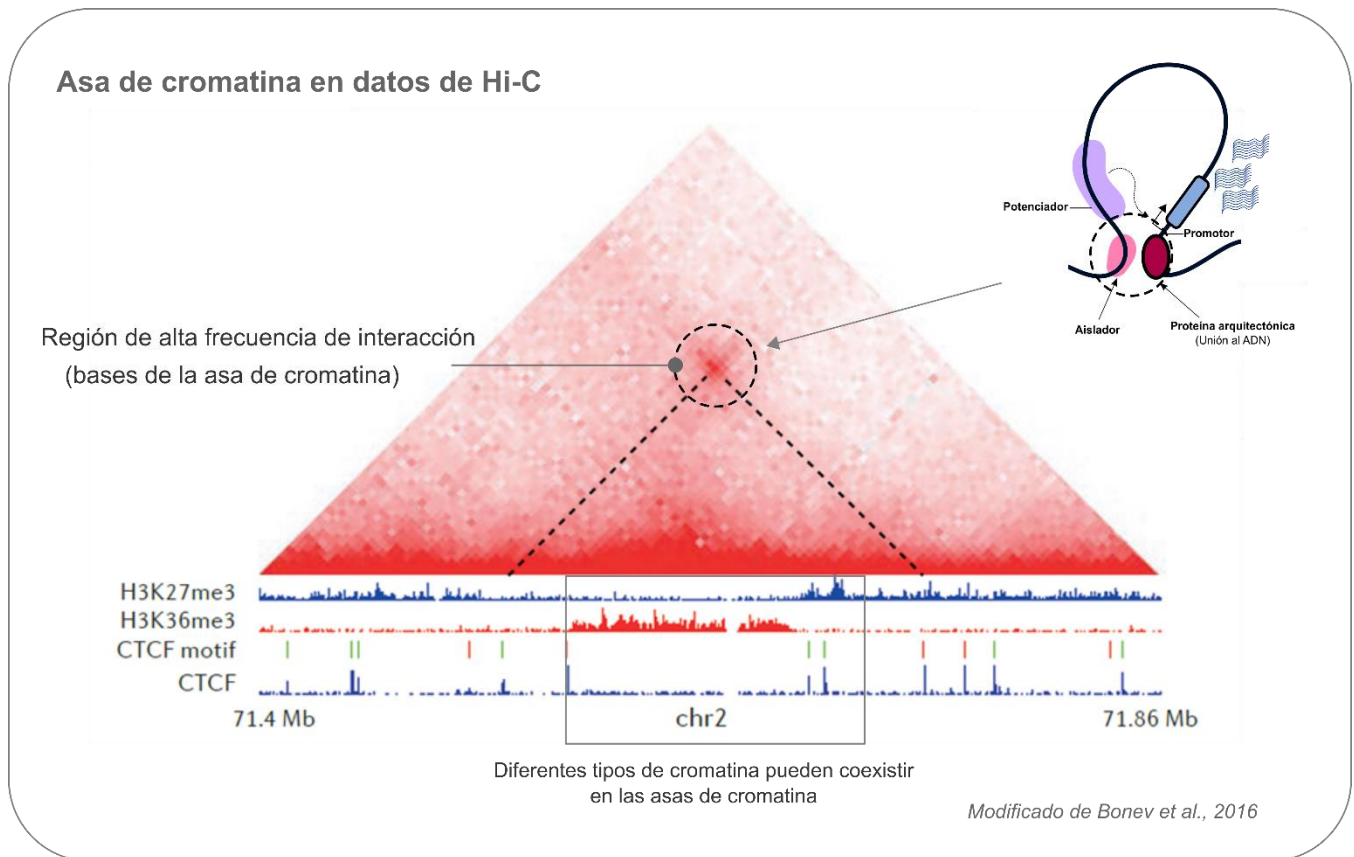


Figura 8. Representación de una asa de cromatina en mapas de calor derivados de experimentos de Hi-C.

Dominios Topológicamente Asociados

También conocidos como Dominios Topológicos o TADs, por sus siglas en inglés, corresponden a regiones de los cromosomas que presentan una alta frecuencia de interacción entre las secuencias al interior del dominio, y una menor frecuencia de interacción con las secuencias adyacentes^{19,20,27,28} (Figura 9). Los TADs son un tipo de dominio que fue identificado a partir de experimentos de Hi-C y se identifican en las matrices de interacción como cuadrados continuos de señal sobre la diagonal²⁰. En mamíferos, el tamaño de las TADs abarca desde algunas decenas de kilobases (~40

kb) hasta el orden de las megabases (3 Mb) siendo el tamaño promedio de ~185 kb²⁶ mientras que en *Drosophila* el tamaño de los Dominios Topológicos se encuentra en el rango de las decenas de kilobases y probablemente refleja el menor tamaño del genoma en comparación con los vertebrados^{25,27-29}. En ambos casos, los Dominios Topológicos tienden a estar asociados con secuencias que presentan un mismo tipo de cromatina, por ejemplo, eucromatina, heterocromatina constitutiva o facultativa^{25,28}. Sin embargo, diferentes estados cromatínicos pueden co-existir en el mismo Dominio Topológico.

Los Dominios Topológicos se encuentran separados por secuencias genómicas conocidas como fronteras. Estas regiones representan sitios en los cuales se aprecia un cambio en la frecuencia de interacción en los mapas de Hi-C y corresponden a sitios de accesibilidad a la cromatina reconocidos por proteínas arquitectónicas como CTCF y Cohesinas²². Adicionalmente, las fronteras de Dominios Topológicos coinciden con promotores de genes constitutivos, sitios de unión a la ARN Pol II y marcas de histonas asociadas a actividad transcripcional como lo son la H3K4me1, me3 y la H3K27ac, así como con la presencia de elementos repetidos.

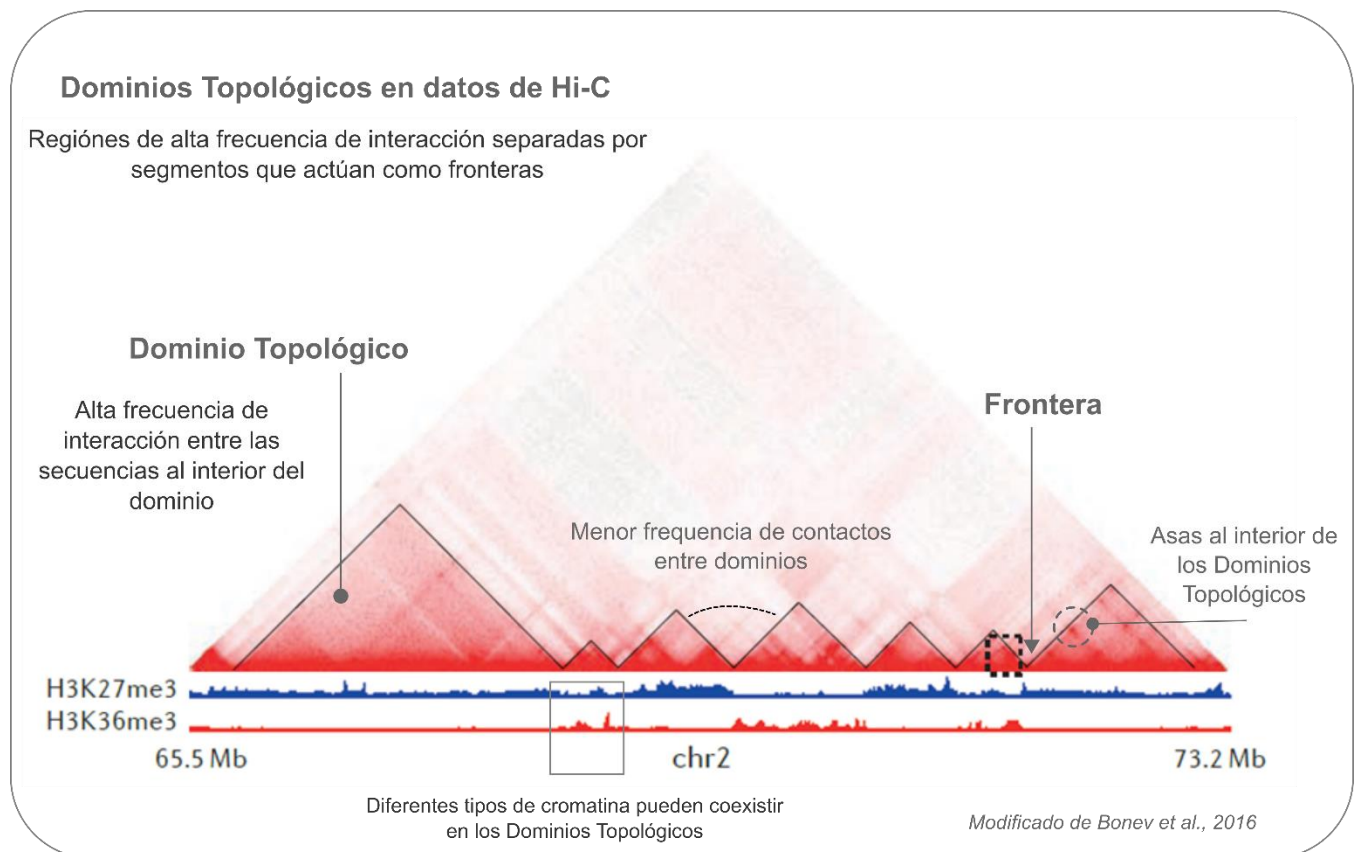


Figura 9. Representación de Dominios Topológicos en datos de Hi-C

La presencia de Dominios Topológicos se ha detectado en múltiples organismos, desde levadura hasta humano y su presencia sugiere que son una característica fundamental de la organización del genoma³⁰. Los Dominios Topológicos parecen estar conservados entre tipos celulares y entre especies aunque ambas observaciones dependen fuertemente de los métodos computacionales empleados para identificar Dominios Topológicos y de la extensión de las regiones homólogas entre especies (Figura 10).

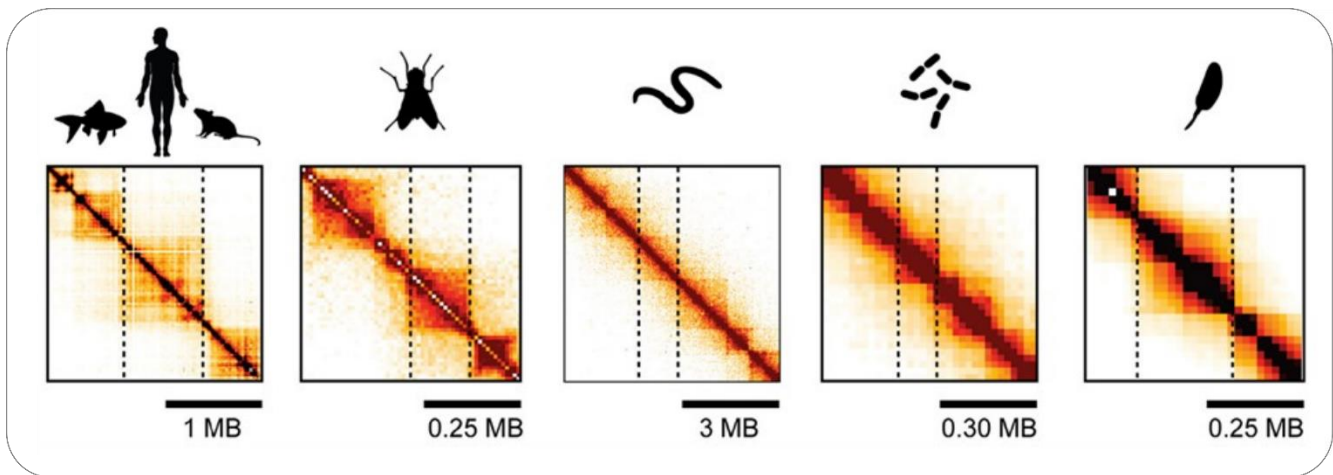


Figura 10. Los Dominios Topológicos son detectados en diferentes especies.

Se ha propuesto que los Dominios Topológicos representan no tan solo unidades arquitectónicas de la cromatina sino también unidades funcionales que pueden definir regiones de regulación transcripcional. Esta propuesta se basa en diversas observaciones: los genes al interior de Dominios Topológicos tienden a ser co-regulados durante la diferenciación celular^{19,31}, una observación que se sostiene principalmente en vertebrados y que correlaciona en parte con un mayor tamaño de los Dominios Topológicos. Por otra parte, los contactos físicos entre elementos enhancer y promotores suele ocurrir al interior de los Dominios Topológicos³². Además, la actividad de genes reporteros insertados en el genoma de mamíferos es influida por enhancers contenidos al interior de Dominios Topológicos³³, un fenómeno muy similar al reportado en experimentos realizados en *Drosophila* en la década de 1980. Finalmente, la perturbación de la estructura de los Dominios Topológicos por manipulación de sus fronteras puede resultar en el establecimiento de contactos ectópicos entre elementos regulatorios que afectan la transcripción de genes y pueden tener importantes implicaciones en el desarrollo de los organismos o papeles centrales en el desarrollo de patologías como el cáncer³⁴⁻³⁶.

Compartimentos

Los compartimentos corresponden a segmentos de la cromatina que tienden a asociarse en función del tipo de modificaciones post-traduccionales de las histonas que conforman a sus nucleosomas. De forma general los compartimentos pueden ser clasificados en compartimentos tipo A y B^{18,25,26} (Figura 11). Los compartimentos tipo A se caracterizan por presentar un tipo de cromatina accesible, con marcas de histonas asociadas a actividad transcripcional y una mayor tasa transcripcional que las regiones que conforman los compartimentos B los cuales presentan una cromatina más compacta y marcas de histonas asociadas a silenciamiento transcripcional. Citológicamente, los compartimentos A y B corresponden a la eucromatina y la heterocromatina, respectivamente. A diferencia de los Dominios Topológicos, los compartimentos tienen una amplia distribución de tamaños, pueden estar contenidos dentro de Dominios Topológicos o bien contenerlos. Además, los compartimentos tienden a establecer contactos de largo alcance entre secuencias que pertenecen a diferentes Dominios Topológicos tanto en el mismo cromosoma como con diferentes cromosomas. En los mapas de Hi-C, los compartimentos suelen ser identificados como regiones de alta frecuencia de interacción fuera de la diagonal.

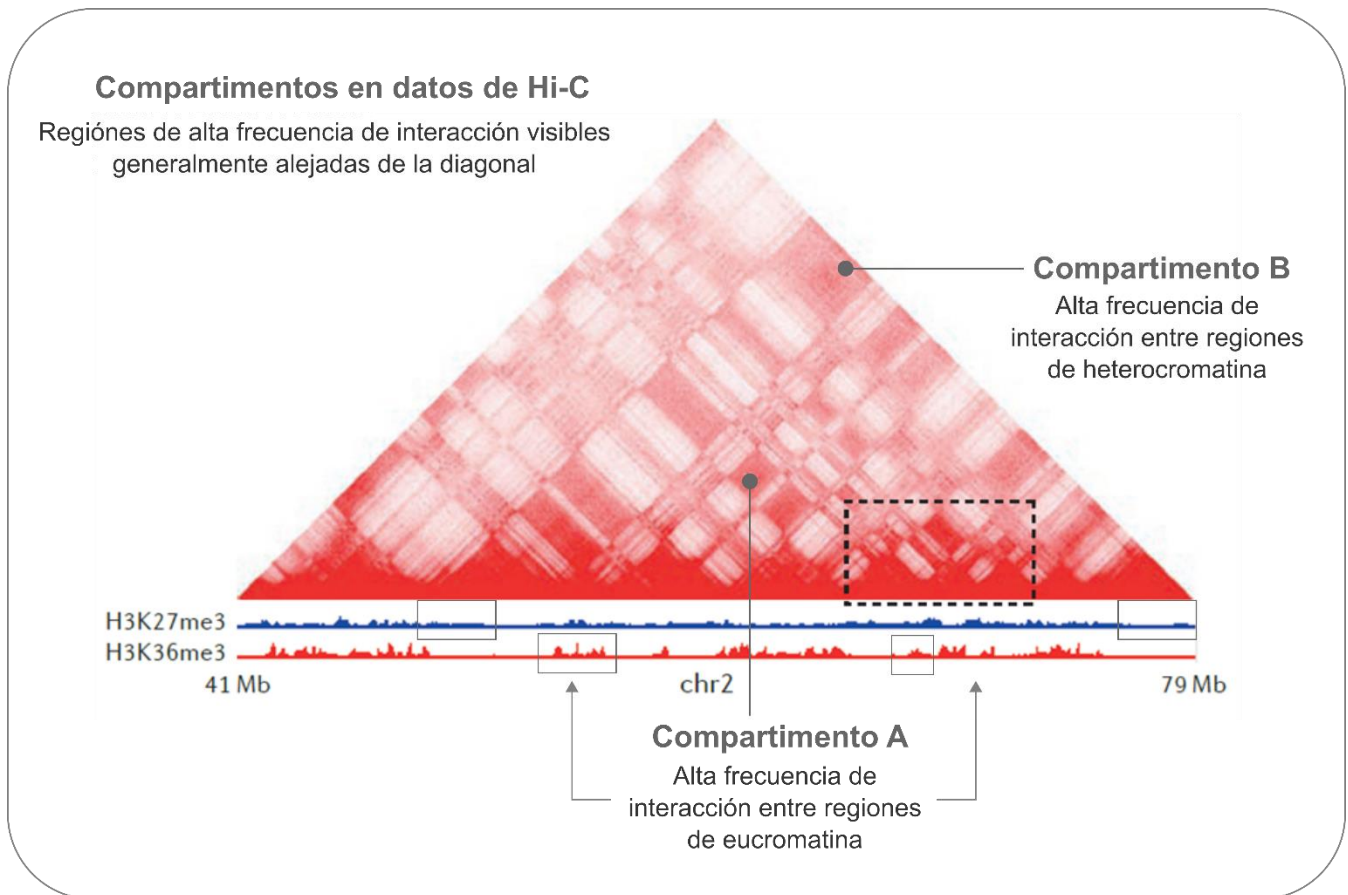


Figura 11. Representación de compartimentos en datos de Hi-C.

Territorios cromosomales

Los territorios cromosomales representan el mayor nivel de organización de cromosomas en interfase. Los territorios cromosomales corresponden a sitios discretos de ocupación del espacio nuclear por cromosomas y presentan una baja frecuencia de interacciones en *trans* con cromosomas adyacentes (Figura 12). Los territorios cromosomales pueden conservar su ubicación espacial hasta cierto grado entre células durante la división celular sin embargo son variables entre tipos celulares. La existencia de territorios cromosomales fue inicialmente sugerida por experimentos de FISH y posteriormente validada por experimentos de Hi-C en donde se observa que los contactos entre loci de un mismo cromosoma son mucho más frecuentes que los observados entre loci de diferentes cromosomas.

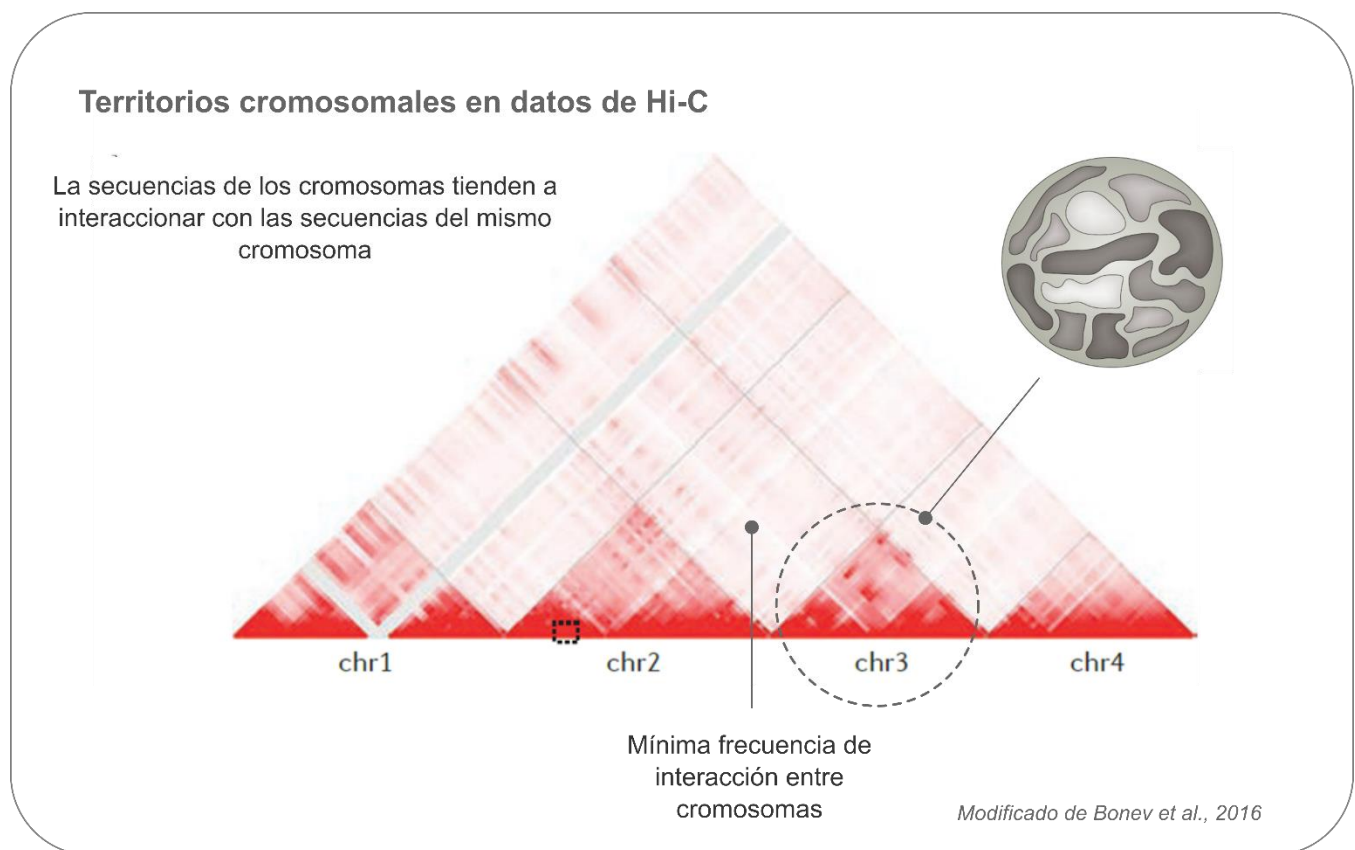


Figura 12. Representación de Territorios cromosomales en datos de Hi-C.

Por lo tanto, podemos pensar en un cromosoma en interfase, como una entidad que presenta diversos niveles de organización tridimensional. Un cromosoma contiene una secuencia finita de nucleótidos en una molécula de ADN. Esta molécula de ADN se asocia con nucleosomas, constituyendo a la fibra de cromatina. Los nucleosomas se agregan entre ellos formando 'clutches' y la fibra de cromatina puede plegarse formando asas de cromatina. Las asas de cromatina pueden agruparse en regiones conocidas como Dominios Topológicos que representan regiones de los cromosomas con una elevada frecuencia de interacciones físicas y que están separados por regiones genómicas conocidas como fronteras de los Dominios Topológicos. Adicionalmente, los nucleosomas que forman la fibra de cromatina pueden ser sujetos de modificaciones post-traduccionales que influyen en la propensión de regiones de la cromatina con perfiles de marcas de histonas similares a agregarse entre sí en el espacio nuclear, constituyendo lo que conocemos como compartimentos y que suelen correlacionar con diferencias en el estado transcripcional.

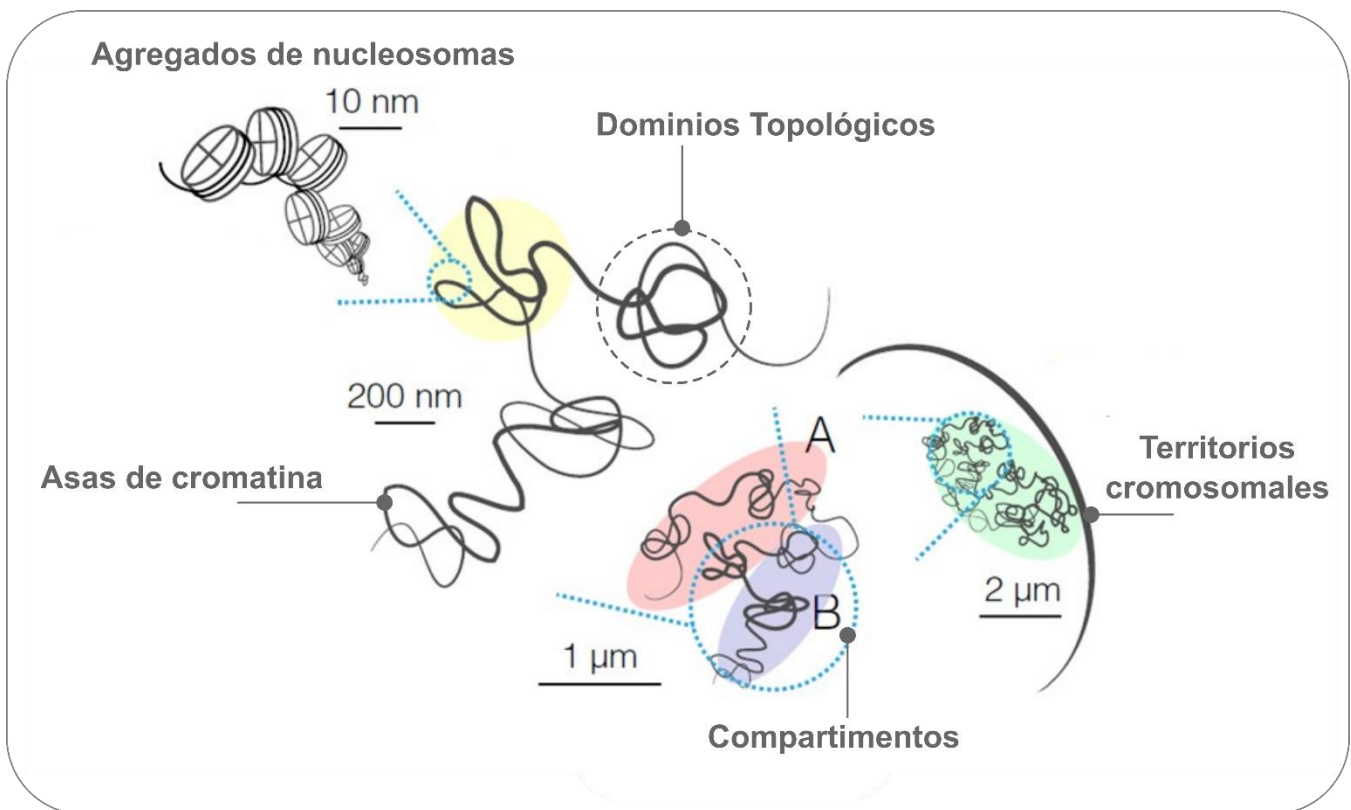


Figura 13. Niveles de organización de los cromosomas en interfase.

Capítulo 3.

Mecanismos de organización 3D del genoma

Vertebrados

Una de las preguntas más importante en el campo de la biología de la cromatina es ¿cómo las características estructurales del genoma se establecen y se mantienen? Actualmente sabemos que la organización tridimensional del genoma en vertebrados depende de una compleja interacción entre la actividad de proteínas arquitectónicas, la presencia de modificaciones post-traduccionales de histonas y de la actividad transcripcional. En esta sección presentaré nuestro entendimiento actual de los procesos moleculares que controlan la organización del genoma en vertebrados.

Proteínas Arquitectónicas

Definimos como Proteínas Arquitectónicas a aquellas proteínas que se encuentran asociadas a las fronteras de los Dominios Topológicos. Esta definición propuesta por Van Bortle y Corces en 2014³⁷ responde a la necesidad de un término sombrilla para hacer referencia a la gran diversidad de proteínas que por experimentos de ChIP-seq se han visto asociadas a regiones genómicas que coinciden con las fronteras de Dominios Topológicos identificados por Hi-C. Las Proteínas Arquitectónicas pueden ser proteínas de unión al ADN, como los factores de transcripción, o bien proteínas asociadas a la cromatina sin reconocimiento específico de secuencia.

De entre todas las Proteínas Arquitectónicas estudiadas hasta la fecha, CTCF es la mejor caracterizada²³. CTCF fue inicialmente caracterizada como una proteína asociada a secuencias *insulator* capaz de restringir contactos físicos entre enhancers y promotores tanto en ensayos transitorios de genes reportero como en su contexto cromatínico. Es una proteína de unión al ADN, con 11 dedos de Zn, altamente conservada entre metazoarios bilaterios, con la excepción notable de una gran variabilidad a nivel de secuencia en sus extremos amino y carboxilo terminal. La expresión de CTCF es constitutiva, sin embargo existen diferencias importantes en el nivel de transcrito y proteína detectados entre tipos celulares. Por ejemplo, durante el desarrollo embrionario, el transcrito de CTCF se ve progresivamente enriquecido en las células del Sistema Nervioso y es particularmente abundante en neuronas. La presencia de CTCF es fundamental para el desarrollo embrionario y la vida post-natal ya que la eliminación del gen resulta en letalidad²³.

Como Proteína Arquitectónica CTCF se localiza en las fronteras de Dominios Topológicos y esta tendencia es consistente en todos los datos de experimentos de Hi-C generados hasta la fecha en vertebrados^{19,26,31,32}. La presencia de CTCF en las fronteras de dominios es consistente con las observaciones de experimentos de bloqueo de enhancer sugiriendo que CTCF es una proteína que influye en la modulación de contactos cromosómicos.

CTCF es fundamental para la organización del genoma en Dominios Topológicos. La eliminación de CTCF en células de ratón provoca la pérdida de Dominios Topológicos y un incremento dramático en los contactos intracromosomales lo cual sugiere una pérdida global de aislamiento³⁸. De forma interesante, la pérdida de CTCF no afecta la organización del genoma nivel de compartimentos, lo cual sugiere que existen al menos dos mecanismos de organización del genoma. En mamíferos, datos de ultra resolución de Hi-C han revelado que las fronteras de Dominios Topológicos interactúan entre sí a través de sitios de unión a CTCF orientados de manera convergente^{26,36}. En este sentido, la eliminación, mutación o inversión de sitios de CTCF en las fronteras de dominios resulta en la pérdida de la frontera y defectos en aislamiento.

Además de CTCF, otras proteínas de unión al ADN se han caracterizado como Proteínas Arquitectónicas. Entre ellas destacan YY1 y la proteína de dedos de Zinc ZNF143. Mientras que ZNF143 también co-localiza con fronteras de dominios, YY1 se localiza frecuentemente en enhancers y promotores que establecen contactos de larga distancia al interior de Dominios Topológicos^{39,40}. La eliminación de YY1 resulta en la pérdida de contactos regulatorios.

Además de CTCF el complejo de Cohesina es fundamental para la organización del genoma en Dominios Topológicos²². La cohesina es un complejo de proteínas en forma de anillo que es importante para la estabilidad de los cromosomas durante la división celular, la segregación cromosomal y la reparación del ADN. En cromosomas en interfase, componentes del complejo de Cohesinas como Rad21 se unen a miles de sitios en el genoma y su unión está enriquecida en las fronteras de Dominios Topológicos. Las cohesinas interactúan físicamente con CTCF a través de su extremo N-terminal y son parte de un mecanismo de extrusión de cromatina que se ha propuesto es responsable de la formación de Dominios Topológicos y asas de cromatina⁴¹ (vea la sección Mecanismos: Extrusión de Cromatina). La eliminación de Cohesinas en células post-mitóticas resulta en la pérdida de todas las asas de la cromatina y Dominios Topológicos^{42,43}. A diferencia de lo observado con la eliminación global de CTCF, la pérdida de cohesinas resulta en la estabilización de los compartimentos. Sin embargo, en ambos casos es claro que la formación de asas de cromatina y Dominios Topológicos es dependiente de CTCF y Cohesinas pero la formación de compartimentos es independiente de estas proteínas.

Transcripción y marcas post-traduccionales de histonas

La presencia de promotores de genes constitutivos, ARN Pol II y de marcas de histonas asociadas a eucromatina en las fronteras de Dominios Topológicos sugiere un papel para la transcripción o bien para proteínas asociadas a la maquinaria basal de la transcripción en la partición en dominio de los cromosomas, sin embargo, la evidencia en mamíferos es controversial. La inhibición de la transcripción en células de ratón tiene efectos mínimos en los Dominios Topológicos. En espermatozoides la organización del genoma en Dominios Topológicos y Compartimentos es claramente identificable pese a que estas células presentan una reducida tasa transcripcional, sin embargo, las fronteras de dominios se encuentran ocupadas por CTCF, Cohesina y ARN Pol II así como por marcas de histonas como H3K4me1, ³⁴⁴. La inhibición de la transcripción antes de la activación cigótica del genoma (ACG) y por lo tanto de la inducción global a la transcripción tiene igualmente efectos reducidos y en el pez cebra los Dominios Topológicos son identificables incluso antes de la ACG⁴⁵. Finalmente, durante la diferenciación neuronal en ratón se observa la ganancia de fronteras de Dominios Topológicos en promotores que son activados durante la diferenciación sugiriendo que la transcripción tiene un papel importante en la formación de fronteras, sin embargo, la inducción de transcripción por el sistema CRISPR-Cas9 acoplado a un potente activador de la transcripción (VP64) no es suficiente para crear una frontera de dominio⁴⁶. Por lo tanto el papel de la transcripción en la formación de dominios topológicos es controversial. En contraste, evidencia creciente sugiere un papel importante para las marcas post-traduccionales de histonas o proteínas asociadas a ellas en la segregación de la cromatina en Compartimentos por un mecanismo denominado Separación de Fases (vea la sección Mecanismos: Separación de fases), sin embargo, la evidencia experimental es indirecta debido a los efectos tóxicos de eliminar la maquinaria enzimática responsable de establecer las marcas post-traduccionales de histonas.

Por lo tanto, evidencia experimental directa sugiere que en vertebrados existen dos mecanismos independientes implicados en la organización tridimensional del genoma en Dominios Topológicos y Compartimentos. Uno dependiente de la actividad de CTCF y el complejo Cohesina y otro que es independiente de estas proteínas. La actividad de CTCF y Cohesinas es necesaria para formar Dominios Topológicos principalmente mediante la formación de asas de cromatina que acercan en el espacio nuclear a sitios de CTCF en orientación convergente. Mientras que un segundo mecanismo es responsable de la segregación de la cromatina en compartimentos y esto ha sido evidenciado principalmente por experimentos en los que se elimina a CTCF y Cohesinas. De forma importante estos hallazgos sugieren que la organización de la cromatina no es jerárquica ni interdependiente y mas bien responde a mecanismos independientes.

Mecanismos I: Extrusión de cromatina

Actualmente el mecanismo más aceptado para explicar la formación de Dominios Topológicos por la acción de CTCF y Cohesinas es el denominado Mecanismo de Extrusión de la cromatina (Figura 14). En este modelo, la cohesina es capaz de generar asas de cromatina por un mecanismo de extrusión^{41,47}. El anillo de cohesina se asocia a una fibra de cromatina y comienza a extruir la fibra de cromatina. La extrusión de la fibra continúa hasta encontrar dos sitios ocupados por CTCF cuyo motivo de unión de encuentra en orientación convergente. En esta orientación el extremo N-terminal de CTCF establece contactos físicos con el anillo de Cohesinas impidiendo que continúe el proceso de extrusión. En este modelo, CTCF actúa como un obstáculo para la procesividad de la Cohesina, confinando el proceso de extrusión y definiendo así lo que observamos en experimentos de Hi-C como fronteras que interaccionan físicamente. Bajo este modelo, las Dominios Topológicos están conformados por múltiples asas de cromatina formadas por la extrusión de la cohesina hasta sitios de CTCF que confinan su actividad. Este modelo tiene amplio soporte de modelado computacional a partir de datos de ultra resolución de Hi-C además de resultados de experimentos en los cuales se elimina a CTCF y Cohesina o se invierten los sitios de unión a CTCF³⁶.

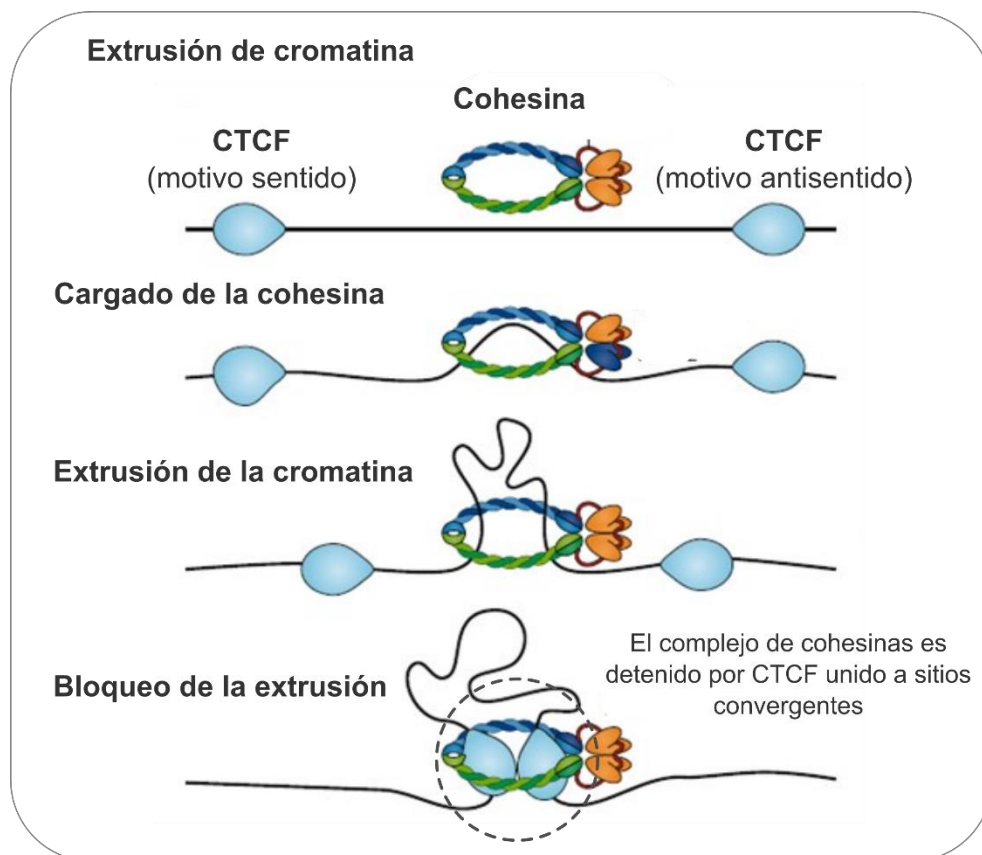


Figura 14. Mecanismo de extrusión mediado por cohesina y CTCF.

Mecanismos II: Separación de fases

Los resultados de experimentos de eliminación de CTCF y Cohesinas demuestran que los compartimentos no se afectan en ausencia de estas proteínas, lo cual sugiere que se forman por al menos un mecanismo adicional e independiente del proceso de extrusión de la cromatina. Diversas líneas de evidencia sugieren que la cromatina en solución tiende a agregarse en micelas dependiente del estado de modificaciones post-traduccionales de histonas o de proteínas asociadas a la cromatina como HP1⁴⁸⁻⁵¹. En particular, existe evidencia sólida que demuestra una tendencia de regiones de heterocromatina a formar micelas en solución⁵². De tal forma que se ha propuesto que la cromatina puede segregarse de forma espontánea en solución en agregados de eucromatina y heterocromatina lo cual se conoce como separación de fases. En este sentido, el proceso de extrusión podría representar un mecanismo activo mediante el cual se traen en proximidad espacial regiones de la cromatina independientemente de su perfil de marcas de histonas. Esto está soportado por dos evidencias indirectas. Por una parte, los Dominios Topológicos pueden contener segmentos de eucromatina y heterocromatina, lo cual es una observación interesante dada la tendencia natural de esos tipos de cromatina a segregarse mutuamente y sugiere que los Dominios Topológicos favorecen la co-existencia de segmentos de eucromatina y heterocromatina. Por otra parte, en ausencia de cohesina, los compartimentos incrementan su frecuencia de interacción, es decir, la segregación entre eucromatina y heterocromatina es más frecuente^{42,43}. La presencia de compartimentos en otros eucariontes que no poseen CTCF, como nematodos, levaduras y plantas sugiere fuertemente que la compartimentalización de la cromatina es un mecanismo basal de organización del genoma²⁵.

Por lo tanto, en vertebrados, la cromatina se organiza al interior del espacio nuclear mediante dos mecanismos. Uno de extrusión de cromatina dependiente de CTCF y Cohesina y que es responsable de la formación de asas de cromatina y Dominios Topológicos y otro que refleja la segregación de la eucromatina y heterocromatina en solución.

Capítulo 4.

Mecanismos de organización 3D del genoma

D. melanogaster

Ochenta años después de la publicación de Müller sobre el efecto de variegación por posición, y a la par de la publicación de los primeros mapas de Hi-C de alta resolución en humano y ratón, los laboratorios de Víctor Corces y de Giacomo Cavalli publicaron de forma independiente los primeros mapas de la organización tridimensional del genoma de *Drosophila*. Los análisis de esos mapas sugerían la presencia de Dominios Topológicos y Compartimentos, similar a lo observado en mamíferos. Sin embargo, pronto fue evidente que la organización tridimensional del genoma en la mosca presentaba diferencias notables con vertebrados y sugería la existencia de mecanismos alternativos de estructuración del genoma al interior del núcleo.

El genoma se organiza en Dominios Topológicos

Los primeros mapas de Hi-C generados para *Drosophila melanogaster* a 10 kb de resolución mostraban la organización de los cromosomas en Dominios Topológicos una observación relevante que sugería que la organización de los cromosomas en regiones de alta frecuencia de interacción era una propiedad conservada al menos entre animales^{27,28}. Los primeros mapas de Hi-C en *Drosophila* sugerían que al igual que en mamíferos, los Dominios Topológicos estaban conservados entre tipos celulares y contenían genes co-regulados. Una observación importante fue que cada Dominio Topológico correlacionaba fuertemente con un estado cromatínico específico, es decir, a diferencia de mamíferos, los Dominios Topológicos en *Drosophila* contenían generalmente cromatina de un mismo tipo, eucromatina o heterocromatina, y por lo tanto se correspondían con estados cromatínicos⁵³. Otra observación importante derivada del análisis de los primeros mapas de Hi-C en *Drosophila* fue la identificación de fronteras y su relación con Proteínas Arquitectónicas⁵⁴. Al igual que en vertebrados, las fronteras de dominio también correlacionaban con la presencia de marcas de histonas asociadas a eucromatina, promotores de genes constitutivos y con secuencias repetidas derivadas de transposones.

Además, datos de Hi-C derivados de cromosomas politénicos demostraron que lo que se observa como bandas en cromosomas politénicos corresponden a Dominios Topológicos y las interbandas a sus fronteras dando por vez primera un ejemplo de

correspondencia directa entre la estructura citológica observada en cromosomas y una caracterización molecular de dicha organización⁵⁵.

Las Proteínas Arquitectónicas se asocian a fronteras de Dominios Topológicos

A diferencia de mamíferos, en *Drosophila* existen más de diez Proteínas Arquitectónicas de unión al ADN caracterizadas hasta la fecha, entre ellas destacan BEAF-32, Su(Hw) y dCTCF. Todas ellas son capaces de bloquear la actividad estimuladora de enhancers con promotores en ensayos de genes reporteros y por lo tanto también se denominan proteínas asociadas a *insulators*⁵⁶. Con excepción del homólogo para CTCF (dCTCF) todas las otras proteínas arquitectónicas son específicas del género de *Drosophila*.

La actividad *insulator* de las Proteínas Arquitectónicas depende de su asociación física con la proteína CP190⁵⁷. A diferencia de lo observado en mamíferos, en *Drosophila* no existe evidencia de la asociación de Proteínas Arquitectónicas con la Cohesina. Esto es sorprendente debido a la presencia de CTCF en *Drosophila*, sin embargo, estudios recientes han demostrado que la capacidad de interactuar físicamente con la Cohesina es una innovación de vertebrados^{58,59}. La proteína CP190 es capaz de mediar contactos de larga distancia a partir de la homodimerización con otros sitios del genoma ocupados por dicha proteína y estudios usando genes reporteros sugieren fuertemente la capacidad de sitios de unión a Proteínas Arquitectónicas en la formación de contactos de larga distancia⁶⁰. Además, CP190 se encuentra asociado a la gran mayoría de las fronteras de Dominios Topológicos en *Drosophila*.

En 2014, el grupo de Victor Corces demostró que las fronteras de Dominios Topológicos identificadas en mapas de Hi-C a una resolución de 10 kb mostraban diferencias importantes con respecto al número de Proteínas Arquitectónicas asociadas³⁷. En concreto, se observó que las fronteras que presentaban mayor capacidad de aislamiento estaban ocupadas por al menos 5 Proteínas Arquitectónicas y consistentemente presentaban la ocupación de CP190. Además, ensayos de bloqueo de enhancer realizados en moscas demostró que aquellas fronteras con ocupación de múltiples Proteínas Arquitectónicas se comportaban como *insulators* fuertes³⁷. Se propuso entonces que la fuerza de una frontera dependía del número y no del tipo de Proteína Arquitectónica asociada. En ese entonces la idea predominante era que los Dominios Topológicos tanto en *Drosophila* como en vertebrados surgían como consecuencia de la actividad de las fronteras de los dominios y que estas eran ocupadas por Proteínas Arquitectónicas de diferente naturaleza.

Los Dominios Topológicos correlacionan con la actividad transcripcional

La publicación de mapas de Hi-C a resolución de 500 pb en el año 2017 demostró diferencias fundamentales en la organización tridimensional del genoma en *Drosophila*²⁵. La observación más importante de estos estudios fue que pese a la organización de los cromosomas en Dominios Topológicos, las fronteras de éstos no establecían contactos físicos de alta frecuencia de interacción como lo observado en mamíferos, es decir, las fronteras de Dominios Topológicos en *Drosophila* no están formados por asas de cromatina. Además, fue evidente que los Dominios Topológicos tenían un tamaño mucho menor que sus contrapartes en mamíferos, con un tamaño promedio de 10 kb, un orden de magnitud menor que en humano. Además, los Dominios Topológicos en *Drosophila* no contenían múltiples genes en su interior y más bien cada Dominio correlacionaba con la presencia de una unidad transcripcional, surgiendo el concepto de Dominio de Gene²⁵. A este nivel de resolución, el grupo de Victor Corces demostró una correlación casi perfecta entre la presencia de Dominios Topológicos y cambios en el estado de la cromatina y la transcripción no solo en *Drosophila*, sino también en ratón, humano y en plantas como *Arabidopsis*, sugiriendo que lo que observamos como Dominios Topológicos en mapas de Hi-C de ultra-resolución es un reflejo de la actividad transcripcional y del estado epigenético de la cromatina lo cual influye en la compactación de los cromosomas²⁵. Esto fue consistente con el marcado enriquecimiento de la ARN Pol II en las fronteras de los dominios de *Drosophila* y se sugirió que la presencia de Proteínas Arquitectónicas confería solo un nivel modesto de aislamiento que correlacionaba con la presencia de CP190^{25,61}. En ese entonces Victor Corces propuso que lo que observamos como fronteras de Dominios Topológicos en realidad son zonas en las que ocurre un cambio o transición en la transcripción y/o marcas de histonas.

En ese mismo año, el grupo de Juan Manuel Vaquerizas caracterizó la dinámica de organización de cromosomas en etapas tempranas del desarrollo embrionario de *Drosophila melanogaster*⁶². Durante el desarrollo embrionario ocurre la activación del genoma del cigoto hacia el ciclo nuclear 14 y esto correlaciona con el reclutamiento global de ARN Pol II por acción de factores transcripcionales pioneros como Zelda⁶³. Datos de Hi-C de alta resolución mostraron que antes de la activación del genoma del cigoto los cromosomas de *Drosophila* no presentaban organización en Dominios Topológicos, una observación muy relevante que sugería que la organización de los cromosomas se establece *de novo* durante el desarrollo. Concomitante con la activación transcripcional del genoma, se observó el surgimiento de Dominios Topológicos, lo cual sugería fuertemente que era la transcripción la responsable de organizar los cromosomas en dominios en *Drosophila*. Sin embargo, la inhibición de la transcripción antes del ciclo nuclear 14 no evitó la formación de Dominios Topológicos consistente con lo observado en mamíferos y sugería que mecanismos alternativos eran responsables de la organización de los cromosomas en Dominios Topológicos. La

inactivación de Zelda y por lo tanto defectos en el reclutamiento de ARN Pol II correlacionaron con la pérdida de fronteras y la fusión de dominios topológicos adyacentes, lo cual sugería que no era la transcripción, si no la presencia de la ARN Pol II lo que influía en la actividad de las fronteras y por lo tanto en la formación de Dominios Topológicos. Sin embargo, la inhibición global de la transcripción, o la eliminación de proteínas regulatorias como Zelda, complican la interpretación de los resultados obtenidos puesto que es difícil descartar efectos colaterales que tengan un impacto global en la organización del genoma.

Capítulo 5.

Antecedentes directos

La organización de la cromatina al interior del núcleo es un proceso regulado con implicaciones funcionales importantes para el mantenimiento y el control de la información genética. En la última década y gracias al uso de técnicas como Hi-C ha sido posible caracterizar la compleja organización espacial de la cromatina al interior del espacio nuclear en múltiples organismos. A partir de estos experimentos es claro que los cromosomas en interfase se estructuran en dominios de cromatina que presentan una elevada frecuencia de interacción. Estos dominios conocidos como Dominios Topológicos representan unidades de organización de los cromosomas.

En vertebrados los Dominios Topológicos se forman en parte a través de un proceso de extrusión mediado por la actividad del complejo de Cohesinas y regulado por la Proteína Arquitectónica CTCF. La extrusión de la fibra de cromatina es responsable de traer en proximidad espacial secuencias genéticas que se encuentran distales en el genoma lineal independiente de su estado epigenético, lo cual sugiere que podría tener implicaciones importantes para la regulación transcripcional al poner en proximidad espacial elementos regulatorios distales.

En *Drosophila melanogaster* los cromosomas también se organizan en Dominios Topológicos, sin embargo, existe un gran debate sobre el o los mecanismos responsables de su formación. En este contexto, se ha planteado que la transcripción, la presencia de ARN Pol II y la ocupación de Proteínas Arquitectónicas influyen en la formación de dominios. En particular, se ha propuesto que los Dominios Topológicos son una consecuencia de la transcripción del genoma y que las fronteras de los dominios son un simple reflejo de las transiciones entre estados transcripcionales. Sin embargo, múltiples líneas de investigación sugieren una función de aisladores o *insulators* a sitios de unión a Proteínas Arquitectónicas, las cuales se asocian frecuentemente a las fronteras de Dominios Topológicos. Sin embargo, no existe evidencia directa a la fecha que sugiera que las fronteras de dominios tienen una actividad autónoma y directa en la formación de Dominios Topológicos en *Drosophila melanogaster*.

Las Dominios Topológicos pueden afectar la regulación de genes al establecer dominios de interacción que permiten la estimulación específica de promotores por secuencias enhancer localizadas al interior de los dominios. Si bien existe evidencia del papel de los Dominios Topológicos en la regulación transcripcional, reportes recientes en mamíferos y *Drosophila* sugieren que esto no es una generalidad, lo cual ha intensificado el debate sobre la relación entre estructura y función del genoma.

Notch como modelo para estudiar mecanismos de organización tridimensional del genoma

Notch es uno de los genes más estudiados en la historia de la Biología del Desarrollo. Codifica a un receptor transmembranal que es fundamental para la diferenciación neuronal temprana y el control de la proliferación en múltiples organismos⁶⁴. En *Drosophila melanogaster*, *Notch* se encuentra en el cromosoma X y mutaciones nulas provocan letalidad temprana a consecuencia de un defecto en el control de la diferenciación de células epiteliales a neuroblastos. Defectos en la actividad de *Notch* también provocan alteraciones en el desarrollo de órganos sensoriales en *Drosophila*, principalmente se afecta la formación del ojo y la ala a consecuencia de defectos en proliferación y diferenciación de células precursoras de dichos órganos durante la etapa larvaria del desarrollo.

El locus de *Notch* en *Drosophila melanogaster* proporcionó la primera evidencia que liga el control de la expresión génica con la organización de la cromatina. En 1975 se reportó una mutante denominada facet-strawberry [*fa(swb)*] y que fenotípicamente presentaba defectos en la formación del ojo^{65,65,66}. Esta mutante fue caracterizada como una mutante hipomórfica *Notch*. Estudios citológicos y genéticos realizados en la misma época permitieron establecer que la unidad transcripcional de *Notch* se localizaba en su totalidad en la banda 3C6 del cromosoma X y se propuso que elementos de control transcripcional podrían estar localizados hacia el extremo 5' que coincidía con la interbanda 3C6-3C7 río arriba de la banda de *Notch*.

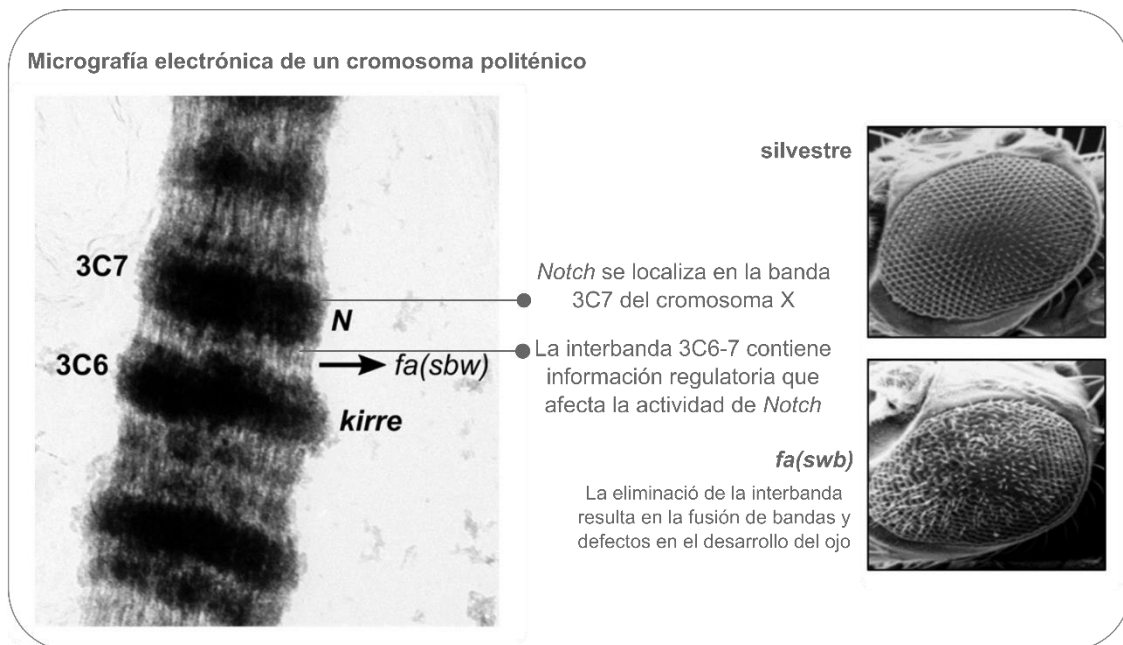


Figura 15. La mutante *fa(swb)* elimina una interbanda río arriba de *Notch*.

De forma interesante, la mutante *fa(swb)* no afectaba la secuencia codificante de *Notch* ni tampoco su promotor y se caracterizaba a nivel citológico por la pérdida de la interbanda 3C6-3C7 que resulta en la fusión de la banda 3C6 que contiene a *Notch*, con la banda río arriba que contiene a diversos genes, entre ellos *white* y *kirre*^{65,66}. En ese entonces se propuso que el efecto de la mutante *fa(swb)* se debía a un efecto de posición que afectaba la transcripción de *Notch* a consecuencia de interacciones regulatorias inespecíficas.

En *Drosophila*, las bandas observadas en cromosomas politénicos corresponden a Dominios Topológicos y las interbandas tienen una buena correspondencia con las fronteras de dichos dominios⁵⁵, lo cual sugiere que el locus de *Notch* es un Dominio Topológico y que su región 5' intergénica corresponde a una frontera de dominio. Por lo tanto, *Notch* podría representar un locus modelo para estudiar la relación entre las fronteras y la organización de los cromosomas en Dominios Topológicos así como su impacto en la transcripción génica.

Experimentos de edición genética realizados en mamíferos fueron fundamentales para comprobar que las fronteras de dominios son elementos genéticos necesarios para la formación de Dominios Topológicos. Experimentos similares en *Drosophila* podrían proveer de información funcional directa sobre la contribución de las fronteras de dominios en la organización tridimensional del genoma. La existencia de una actividad autónoma de fronteras de dominios en *Drosophila melanogaster* sugeriría que la información necesaria para plegar los cromosomas al interior del núcleo está codificada en la secuencia genética y que a lo largo de la evolución diversos mecanismos moleculares han surgido para leer esa información y organizar los cromosomas en Dominios Topológicos.

Capítulo 6.

Propuesta de investigación

Planteamiento del problema

En eucariontes, el genoma se organiza al interior del núcleo en Dominios Topológicos. En *Drosophila melanogaster* los Dominios Topológicos se encuentran separados por fronteras de dominios que corresponden a sitios de accesibilidad a la cromatina ocupados por múltiples Proteínas Arquitectónicas y ARN Pol II, y se piensa son importantes para la organización tridimensional del genoma y la regulación transcripcional. Por otra parte, se ha propuesto que es la transcripción y no las fronteras de dominios el mecanismo responsable de la organización del genoma en Dominios Topológicos.

Hipótesis

Las fronteras de dominios son necesarias para la formación de Dominios Topológicos y por lo tanto para la organización tridimensional del genoma en *Drosophila melanogaster*. La eliminación *in situ* de fronteras de dominios resulta en la fusión de Dominios Topológicos adyacentes y en cambios en la transcripción de los genes al interior de cada dominio.

Objetivos

En la presente investigación estudiamos la contribución de las fronteras de Dominios Topológicos en la organización tridimensional del genoma y su impacto en la expresión de genes empleando al locus de *Notch* de *D. melanogaster* como modelo de estudio. Para esto planteamos realizar eliminaciones precisas de las fronteras de Dominio asociadas al locus *Notch* y evaluar su efecto en la topología del genoma y en la transcripción de genes.

Objetivos metodológicos

1. Establecimiento de un protocolo para la generación de mutantes *in situ* mediante el sistema CRISPR-Cas9 en líneas celulares de *D. melanogaster*.
2. Establecimiento del protocolo de *in nucleus* Hi-C para líneas celulares de *D. melanogaster*.
3. Establecimiento de un protocolo de análisis computacional de datos de *in nucleus* Hi-C.

Objetivos relacionados con la pregunta de investigación

1. Caracterización de la organización tridimensional de *Notch* en la línea celular S2R+ de *D. melanogaster* mediante *in situ* Hi-C.
2. Generación de líneas celulares mutantes para las fronteras de Dominios Topológicos de *Notch* mediante el sistema CRISPR-Cas9.
3. Caracterización de la organización tridimensional de *Notch* en las líneas celulares mutantes para las fronteras de Dominios Topológicos de *Notch* mediante *in situ* Hi-C.
4. Caracterización del estado de la cromatina y reclutamiento de Proteínas Arquitectónicas de *Notch* en las líneas celulares mutantes para las fronteras de Dominios Topológicos de *Notch* mediante qChIP.
5. Caracterización de los efectos transcripcionales en las líneas celulares mutantes para las fronteras de Dominios Topológicos de *Notch* mediante RT-PCR en tiempo real y RNA-seq.

Los resultados asociados a los Objetivos experimentales y analíticos relacionados con la pregunta de investigación se presentan en el Capítulo 7 de esta tesis. Las figuras asociadas a nuestros resultados se muestran al final del Capítulo 7.

Capítulo 7.

Resultados

“How can I help seeing what is in front of my eyes. Two and two are four. Sometimes...sometimes they are five. Sometimes they are three. Sometimes they are all of them at once. You must try harder. Its not easy to be sane”

1984

George Orwell

I. El locus de *Notch* se organiza en dos dominios topológicos en la línea celular embrionaria S2R+

En *D. melanogaster* el locus de *Notch* corresponde a una unidad transcripcional de ~40 kb localizada entre los genes *kirre* y *dnc*. A nivel citológico, *Notch* se localiza en la banda 3C7 del cromosoma X mientras que *kirre* comprende las bandas 3C5-6 y *dnc* a las bandas 3C8-9⁶⁵⁻⁶⁷. Estudios citológicos de cromosomas politénicos han establecido que el locus de *Notch* es la única unidad transcripcional presente en la banda 3C7 del cromosoma X⁶⁷. Además, resultados de estudios de hibridación *in situ* de alta resolución para el locus de *Notch* han demostrado que la banda 3C7 está dividida en dos bandas de diferente peso molecular⁶⁷, sugiriendo una estructuración adicional de la banda 3C7 que contiene a *Notch*.

Estudios genéticos y citológicos han establecido como el límite 5' del locus de *Notch* a la interbanda que separa a la banda 3C6 y 3C7. En moscas mutantes que presentan la eliminación de la región 5' no codificante de *Notch* se observa la pérdida de la interbanda y la fusión de las bandas 3C6 y 3C7, además de defectos en la vía de *Notch* visibles como defectos en la formación del ojo en *D. melanogaster*⁶⁶. Lo anterior sugiere que el extremo 5' de *Notch* codifica para un aislador de cromatina que impacta la organización en bandas de los cromosomas politénicos en *D. melanogaster*.

En *D. melanogaster*, las bandas de cromosomas politénicos se corresponden con Dominios Topológicos (TADs) identificados por Hi-C, mientras que las interbandas se corresponden con las fronteras de los Dominios Topológicos⁵⁵, lo cual sugiere que la banda 3C7 que contiene únicamente al locus de *Notch* podría corresponderse con al menos un Dominio Topológico cuya frontera 5' correspondería con la interbanda detectada entre las bandas 3C6 y 3C7.

Para caracterizar la organización tridimensional del locus de *Notch* realizamos experimentos de *in nucleus* Hi-C en la línea celular embrionaria S2R+ (vea detalles en Métodos). La Inspección visual de los datos de Hi-C a diferentes resoluciones revela la presencia de Dominios Topológicos en células S2R+ consistente con observaciones previas en otras líneas celulares y embriones de *D. melanogaster*^{27,28,62,68-70}. Un mapa de calor o “heatmap” de datos de Hi-C a 5 kb de resolución en la región genómica comprendida entre el locus *kirre* y *dnc* revela la presencia de al menos cuatro Dominios Topológicos (Figura Suplementaria 1). El locus *kirre* de ~400 kb se organiza en dos dominios de contacto de ~200 kb cada uno que denominamos *kirre*-D1 y *kirre*-D2, mientras que a 5kb de resolución se observa que el locus *Notch* corresponde a un Dominio Topológico al igual que el locus *dnc*. La organización en Dominios Topológicos de la región comprendida entre *kirre* y *dnc* presenta una buena correspondencia con la localización de bandas de cromosomas politénicos reportadas para esta región.

Para obtener mayor información sobre la organización tridimensional del locus de *Notch*, generamos mapas de calor empleando nuestros datos de Hi-C a 1kb de resolución. La inspección visual de datos de Hi-C a 1kb de resolución de una región de ~50 kb centrada en el locus de *Notch* demuestra que *Notch* se encuentra organizado en dos Dominios Topológicos en células S2R+ (Figura 16a). Para validar nuestra observación, identificamos computacionalmente Dominios Topológicos en todo el genoma empleando matrices de interacción a 1kb de resolución y, consistente con lo observado en los mapas de interacción, el locus de *Notch* se encuentra particionado en dos Dominios Topológicos (Figura 16a y Figura Suplementaria 1b). El Dominio 1 de *Notch* tiene un tamaño de 31 kb y comprende desde el inicio del gen hasta el exón 6 mientras que el Dominio 2 tiene un tamaño de 9 kb y contiene a los 4 exones restantes. Por lo tanto, nuestros datos de *in nucleus* Hi-C a 1kb de resolución demuestran que *Notch* se organiza en el espacio nuclear en dos Dominios Topológicos de distinto tamaño en células S2R+. Esta conformación tridimensional es consistente con lo observado a nivel citológico para el locus de *Notch*⁶⁷.

II. Las fronteras de los Dominios Topológicos de *Notch* son elementos regulatorios reconocidos por Proteínas Arquitectónicas y ARN Pol II

Nuestro análisis de Dominios Topológicos en células S2R+ a 1kb de resolución revela la existencia de dos fronteras de dominio en el locus de *Notch*. La primera frontera, que denominamos frontera F1 (por Frontera 1), está localizada en la región intergénica entre los genes *kirre* y *Notch* y separa al Dominio 1 de *Notch* del dominio *kirre*-D2 (Figura 16 a). La segunda frontera, que denominamos como frontera F2 (por Frontera 2), está localizada en la región genómica comprendida entre el final del intrón 5 y el inicio del exón 6 de *Notch* y coincide con la transición entre los Dominios 1 y 2 de *Notch* (Figura 16a).

En *D. melanogaster*, las fronteras de Dominios Topológicos corresponden frecuentemente a sitios accesibles a la cromatina y que se encuentran ocupados por diferentes Proteínas Arquitectónicas y por componentes de la maquinaria transcripcional como la ARN Pol II ^{27,37,62}. Como una primera aproximación para caracterizar el estado cromatínico y de ocupación de Proteínas Arquitectónicas en las fronteras del locus de *Notch* analizamos datos de ChIP-seq públicos contra Proteínas Arquitectónicas, ARN Pol II, marcas post-traduccionales de histonas y accesibilidad a la cromatina derivados de células S2/S2R+ (vea detalles en Métodos).

La frontera F1, localizada en la región intergénica 5' de *Notch*, es un sitio accesible a la cromatina, ocupado por la ARN Pol II, nueve Proteínas Arquitectónicas con capacidad de unión al ADN y por los co-factores de Proteínas Arquitectónicas CP190 y Mod(mdg4). Interesantemente, y pese a la presencia de múltiples Proteínas Arquitectónicas en la frontera F1, solo detectamos los motivos de unión para dCTCF y M1BP (FIMO *p*-valor < 0.0001). A nivel de marcas de histonas, observamos el enriquecimiento de marcas post-traduccionales asociadas a cromatina accesible como la H3K4me3, H3K4me1, H3K27ac y H4K16ac y la ausencia de enriquecimiento para la marca post-traduccional asociada a silenciamiento H3K27me3. Además, datos de modENCODE⁷¹ para la línea celular S2 clasifican a parte de la región intergénica de *Notch* como un promotor, en función de las marcas de histonas presentes en esa zona, lo cual es consistente con observaciones previas que sugieren que las fronteras de Dominios Topológicos en *Drosophila* se localizan frecuentemente en regiones promotoras^{27,62}.

La frontera F2 está localizada hacia el final del intrón 5 de *Notch* (Figura 16a). Consistente con lo observado para la frontera F1, esta región se encuentra ocupada por la ARN Pol II y presenta enriquecimiento de las marcas post-traduccionales H3K4me3, H3K4me1, H3K27ac y H4K16ac, sin embargo observamos que este enriquecimiento es mayor que el observado para la frontera F1. Además, y en contraste con lo observado para la frontera F1, la frontera F2 es un sitio de baja ocupación por Proteínas Arquitectónicas, ya que observamos la presencia de solo tres de estas proteínas además de los co-factores de Proteínas Arquitectónicas CP190 y Mod(mdg4). Interesantemente, esta región genómica ha sido previamente descrita como un elemento *enhancer* que controla la transcripción de *Notch* durante el desarrollo larvario⁷² consistente con observaciones previas que sugieren que las fronteras de Dominios Topológicos en *Drosophila* coinciden con elementos tipo *enhancer*^{27,56,68}.

También observamos que en el Dominio 2 de *Notch*, el exón 6 presenta una señal de enriquecimiento focal para la ARN Pol II y un enriquecimiento prominente de las marcas de histonas H3K4me3, H3K4me1 y H3K27ac alrededor del sitio de unión para la ARN Pol II y que se encuentra a ~1kb de la frontera F2 del locus (Figura 16 a y 19a). Estas observaciones provenientes de datos públicos de ChIP-seq sugieren que en el exón 6 de *Notch* podría existir un elemento regulatorio adicional que influya en la transcripción del locus.

Para validar el perfil de enriquecimiento de marcas de histonas y la presencia de ARN Pol II así como de Proteínas Arquitectónicas en el locus de *Notch*, realizamos ensayos de Inmunoprecipitación de la cromatina y qPCR (ChIP-qPCR por sus siglas en inglés) contra las marcas de histonas H3K4me3, H3K27ac, H3K27me, la ARN Pol II pSer2 y pSer5 así como de CTCF. Validamos la unión de CTCF, ARN Pol II (pSer2 y pSer5) así como el enriquecimiento de las marcas de histonas H3K4me3, H3K27ac y H3K27me3 en las fronteras de dominios en *Notch* (Figura 16b). Consistente con lo observado en datos de ChIP-seq públicos, CTCF se asocia de forma exclusiva a la frontera F1 (Figura 16b izquierda). Por otra parte, pese a que de la marca de histonas H3K27ac está enriquecida en ambas fronteras, la frontera F2 muestra un enriquecimiento de hasta 10 veces superior a los niveles en la frontera F1, lo cual es consistente con la presencia de un elemento enhancer en la frontera F2 (Figura 16b derecha).

Validamos la unión de ARN Pol II en el exón 6 a niveles comparables a los identificados para la frontera F1 y observamos que la marca de histonas H3K27ac se encuentra enriquecida a niveles comparables a los caracterizados para la frontera F2 (Figura 16b). Adicionalmente, validamos la presencia de la marca de histonas asociada a promotores H3K4me3 a un nivel de enriquecimiento incluso superior al detectado para la frontera F1. Estos datos sugiere que en el exón 6 de *Notch* existe una región genómica con una estructura de la cromatina similar a la que caracteriza a promotores.

Por lo tanto, los datos presentados sugieren que en células S2R+, las fronteras de Dominios Topológicos detectadas en el locus de *Notch* coinciden con dos elementos regulatorios que presentan enriquecimiento para marcas post-traduccionales de histonas asociadas a eucromatina y son sitios de unión para Proteínas Arquitectónicas y de reclutamiento de ARN Pol II.

III. La organización de *Notch* en dos Dominios Topológicos se establece gradualmente durante el desarrollo embrionario

En *D. melanogaster*, la organización del genoma en Dominios Topológicos surge durante el ciclo nuclear catorce y coincide con la Activación del Genoma del Cigoto^{62,69}. Durante este proceso, la ARN Pol II es reclutada al genoma gracias a factores transcripcionales como Zelda⁶³. La unión de ARN Pol II durante el desarrollo embrionario temprano coincide con el establecimiento de fronteras de Dominios Topológicos y con la activación transcripcional de cientos de genes. Puesto que *Notch* es un gen fundamental para el desarrollo embrionario de *D. melanogaster* evaluamos la dinámica de formación de Dominios Topológicos en el locus de *Notch* durante el desarrollo embrionario temprano analizando datos públicos de *in nucleus* Hi-C⁶².

El análisis de datos de Hi-C durante el desarrollo embrionario temprano en *Drosophila* a 1kb de resolución revela el establecimiento progresivo de dos Dominios Topológicos

en el locus de *Notch* (Figura Suplementaria 2). En el ciclo nuclear 12 (nuclear cycle 12; nc12) no se detectan fronteras de Dominios Topológicos en el locus de *Notch*, consistente con la ausencia de Dominios Topológicos a nivel de genoma completo en ese estadio embrionario, sin embargo, observamos una frontera de dominio en el extremo 5' del locus en el ciclo nuclear 13 (nuclear cycle 13; nc13) mientras que una segunda frontera es detectada entre el intrón 5 y exón 6 durante el ciclo nuclear 14 (nuclear cycle 14; nc14), coincidente con la Activación del Genoma del Cigoto (Figura Suplementaria 2a). Ambas fronteras dividen al locus de *Notch* en dos Dominios Topológicos similar a lo observado en la línea celular S2R+.

El establecimiento de fronteras de Dominios Topológicos durante el desarrollo embrionario de *D. melanogaster* coincide con el reclutamiento de la ARN Pol II y el establecimiento de sitios de accesibilidad a la cromatina⁶². En el caso de *Notch*, observamos que el establecimiento de las fronteras de dominios coincide con la ganancia progresiva de accesibilidad de la cromatina en ambas fronteras, así como la unión de RNA Pol II, TBP y el factor de transcripción Zelda (Figura Suplementaria 2b).

Por lo tanto, nuestros resultados indican que el locus de *Notch* se organiza en dos Dominios Topológicos. Las fronteras de *Notch* presentan accesibilidad a la cromatina, enriquecimiento para marcas de histonas como H3K27ac y están ocupadas por la ARN Pol II y Proteínas Arquitectónicas como dCTCF. Durante el desarrollo embrionario, las fronteras de dominio en *Notch* se establecen incluso antes de la transcripción del locus y su presencia detectada por Hi-C correlaciona con la adquisición de accesibilidad a la cromatina, la unión de factores de transcripción y de la ARN Pol II.

IV. La región intergénica 5' de *Notch* es una frontera de Dominios Topológicos modular

Para evaluar directamente la contribución de la frontera F1 en la organización tridimensional del locus de *Notch* establecimos un protocolo de edición genética basado en el sistema CRISPR-Cas9 en células S2R+ para generar eliminaciones de las fronteras de interés y obtener clones mutantes (vea Métodos).

Puesto que la frontera F1 localiza con la región intergénica 5' de *Notch*, realizamos el diseño de mutantes acotados a dicha región (tamaño de 300 pb). Dividimos la región intergénica 5' de *Notch* en dos segmentos en función de la presencia de los motivos de unión al ADN de las Proteínas Arquitectónicas identificadas en dicha zona (Figura 1d y Figura Suplementaria 3). Denominamos como la Región Proximal al Promotor (RPP) a una secuencia de ~100 pb que contiene el motivo de unión a la Proteína Arquitectónica M1BP y denominamos como la Región Intergénica (RI) a una secuencia de ~200 pb que contiene al sitio de unión a dCTCF. Las regiones genómicas comprendidas por la RPP y la RI son adyacentes y no sobrelapan entre sí. En función de

esta organización de la región 5' intergénica de *Notch* diseñamos cuatro tipos de eliminaciones sobre la frontera F1. Dos de las eliminaciones remueven individualmente a la RPP y la IR, mientras que otra de las eliminaciones remueve tanto a la RPP y la IR y por lo tanto la totalidad de la región intergénica 5' de *Notch* (Figura 1d y Figura Suplementaria 3). Además, diseñamos una eliminación de ~8000 pb que corresponde a la mutante *fa(swb)* que ocasiona la fusión de la banda de *Notch* en cromosomas politénicos^{73,74}. Es importante mencionar que ninguna mutante elimina el sitio de inicio de la transcripción de *Notch*⁷³.

Para generar eliminaciones de segmentos de ADN empleando el sistema CRISPR-Cas9 optamos por una estrategia de dos sgRNAs por región genómica a eliminar, de tal forma que cada guía se posiciona en los extremos de la zona que deseamos eliminar (vea Métodos). Empleando esta estrategia y en función del esquema de mutantes diseñado, generamos cuatro tipos de eliminaciones de la frontera F1 (Figura 1c,d). Aislamos clones mutantes para las cuatro diferentes eliminaciones y cada mutante fue nombrada en función del número de pares de bases eliminados. Para evaluar el efecto de cada mutante en la organización tridimensional de *Notch* realizamos experimentos por duplicado de *in nucleus* Hi-C, así como también caracterizamos el perfil de marcas de histonas, unión de Proteínas Arquitectónicas y reclutamiento de ARN Pol II mediante qChIP y caracterizamos el efecto en la transcripción de *Notch* y genes adyacentes empleando qRT-PCR.

La mutante *5pN-Δ102* se caracteriza por la eliminación de 102 pares de bases (pb) de la región que denominamos como Región Proximal al Promotor (RPP). Esta mutante carece del motivo de unión al ADN de la Proteína Arquitectónica M1BP y los *peak summits* de las proteínas arquitectónicas dCTCF, Su(Hw), Pita y BEAF-32 (Figura 1a,c-d). A nivel topológico, la mutante *5pN-Δ102* se caracteriza por la pérdida significativa de las interacciones físicas entre secuencias al interior del Dominio 1 (*p*-valor 6.87e-9). En particular, observamos una pérdida de interacciones entre los subdominios del Dominio 1, mientras que el Dominio 2 muestra una disminución no significativa de interacciones (Figura 17a). De forma interesante, observamos que la región 5' UTR de *Notch*, que contiene el principal sitio de inicio de la transcripción, muestra una disminución de contactos con secuencias del Dominio 1 y una ganancia de contactos ectópicos con el dominio kirre-D2 que se localiza adyacente y río arriba (Figura Suplementaria 4), sin embargo, la ganancia de contactos ectópicos entre otras secuencias al interior del Dominio 1 con el dominio kirre-D2 no se observa en esta mutante (Figura 17c). Estos datos sugieren que, pese a los efectos topológicos observados en la mutante *5pN-Δ102*, la eliminación de la Región Proximal al Promotor de *Notch* (RPP) no es suficiente para inducir la fusión del Dominio 1 con el dominio río arriba.

La mutante *5pN-Δ183* se caracteriza por una eliminación de 183 pb río arriba de la región proximal al promotor que denominamos como Región Intergénica (RI). Esta mutante carece del sitio de unión al ADN de la Proteína Arquitectónica dCTCF y los

peak summits de las proteínas arquitectónicas dCTCF, lbf1 e lbf2 (Figura 1a,d). A nivel topológico la mutante *5pN-Δ183* no presenta cambios en la organización tridimensional de *Notch* (Figura 17a). Sin embargo, se observa un incremento significativo de contactos ectópicos entre los Dominios Topológicos localizados en el locus *kirre* (*kirre*-D1 y D2) y el Dominio Topológico que comprende al locus *dnc* (*p*-valor $2.2e-16$) (Figura 17b,d). La ganancia de contactos entre los Dominios Topológicos de *kirre* y *dnc* se acompaña con una disminución en los contactos entre las secuencias que conforman a cada Dominio Topológico (Figura 17d), lo cual podría sugerir que, ante la pérdida de la Región Intergénica de *Notch* (RI), secuencias genómicas al interior de los Dominios *kirre*-D1,2 pierden aislamiento y establecen contactos ectópicos con regiones adyacentes (Figura 17d). Por ejemplo, en la mutante *5pN-Δ183* observamos la ganancia de un contacto ectópico de larga distancia entre los Dominios Topológicos *kirre*-D2 y *dnc* (Figura 17d arcos). Este contacto refleja un incremento en la frecuencia de interacción entre secuencias ocupadas por Proteínas Arquitectónicas, CP190 y enriquecidas para marcas de histonas asociadas a enhancers. Además, se observa que las interacciones ectópicas entre los Dominios Topológicos *kirre*-D1,2 y *dnc* se extienden río arriba hasta la frontera 5' del dominio *kirre*-D1 y río abajo hasta la frontera 3' del dominio *dnc* (Figura 17d). Estas fronteras están ocupadas por proteínas como CTCF, CP190 y ARN Pol II. Esto sugiere que ambas fronteras establecen interacciones de larga distancia en la mutante *5pN-Δ183* probablemente a consecuencia de la preferencia de las Proteínas Arquitectónicas en cada frontera a interactuar físicamente en el espacio nuclear⁶⁰. Por lo tanto, la Región Intergénica de *Notch* (RI) es esencial para contener interacciones físicas entre elementos genéticos dentro del locus *kirre*, mientras que es dispensable para mantener la estructuración tridimensional de *Notch*.

Nuestros datos indican que la frontera F1 de *Notch* afecta la organización en Dominios Topológicos de las secuencias adyacentes y que es modular. La Región Proximal al Promotor de *Notch* (RPP) es importante para estructurar el Dominio 1 de *Notch* mientras que la Región Intergénica (RI) es importante para la organización en Dominios Topológicos del locus *kirre*. Si bien esto sugiere que ambas secuencias son importantes para la actividad de la frontera F1, ninguna de estas mutaciones resulta en la fusión de los Dominios Topológicos adyacentes a la frontera F1, un efecto observado en vertebrados y que acompaña a la inactivación de una frontera topológica^{34,35,75}. Puesto que en cada mutante se pierde solo uno de los dos motivos de unión al ADN para las proteínas MB1P (*5pN-Δ102*) y CTCF (*5pN-Δ183*) así como una fracción de los *peak summits* para Proteínas Arquitectónicas, es posible que la fusión de Dominios Topológicos adyacentes a la frontera F1 de *Notch* requiera de la eliminación de todos los motivos de unión al ADN para Proteínas Arquitectónicas. Para evaluar, esta posibilidad, eliminamos la secuencia de ADN comprendida por las mutantes *5pN-Δ102* y *5pN-Δ183*. La pérdida de la Región Proximal al Promotor en conjunto con la Región Intergénica de *Notch* (IR+PPR) (*5pN-Δ343*) resulta en la fusión de dominios de contacto adyacentes a la frontera F1 (Figura 17a,b y e). En particular, observamos la pérdida de interacciones al interior del Dominio 1 de *Notch* (*p*-valor $2.2e-16$) acompañada por la

ganancia de contactos ectópicos con el Dominio Topológico río arriba, kirre-D2 (p -valor $2.2e-16$) (Figura 17b,c y e).

Para confirmar la pérdida de la frontera F1 identificamos computacionalmente fronteras de Dominios Topológicos en todo el genoma usando datos de Hi-C para las mutantes *5pN-Δ102*, *5pN-Δ183* y *5pN-Δ343*. Consistente con lo observado en los mapas de interacción, la frontera F1 no es identificada computacionalmente en datos de Hi-C de la mutante *5pN-Δ343* pero si en las mutantes *5pN-Δ102* y *5pN-Δ183*, lo cual confirma la inactivación de la frontera F1 de *Notch* (Figura 17a, observe las TADs debajo de la representación triangular de los mapas de Hi-C). Además, encontramos que en la mutante *5pN-Δ343* el Dominio Topológico kirre-D2 se extiende río abajo y ahora incluye las 31 kb que originalmente conformaban al Dominio 1 de *Notch*, siendo detectada la frontera F2 de *Notch* como la nueva frontera 3' del dominio extendido kirre-D2 (Figura 17e). Finalmente, observamos un efecto topológico similar cuando eliminamos un segmento de ~800 bp reportado para la mutante *fa(sbw)* y que resulta en la fusión de banda adyacentes en cromosomas politénicos.

Por lo tanto, la frontera F1 de *Notch* es una frontera *bona fide* de Dominios Topológicos. Esta frontera está codificada en ~300 pb que corresponden a la región intergénica 5' de *Notch*. Nuestros datos demuestran que esta frontera está conformada por dos módulos que de manera independiente influyen en la organización 3D de los dominios adyacentes. La remoción de ambos módulos, que contienen los sitios de unión para las Proteínas Arquitectónicas CTCF y M1BP, es necesaria y suficiente para la inactivación de esta frontera y la fusión de los Dominios Topológicos adyacentes.

V. La frontera F1 *Notch* es necesaria para el reclutamiento de Proteínas Arquitectónicas y ARN Pol II

Se ha propuesto que la presencia de proteínas arquitectónicas así como el reclutamiento de ARN Pol II en fronteras de Dominios Topológicos es importante para la organización del genoma en *D. melanogaster*^{29,37,62}. Para correlacionar los efectos topológicos observados en nuestras mutantes con la presencia de Proteínas Arquitectónicas y marcas de histonas, realizamos ChIP en cada una de las cuatro mutantes para la frontera F1 empleando anticuerpos contra CTCF, ARN Pol II (pSer2 y pSer5) y las marca de histonas H3K27ac. Evaluamos el enriquecimiento de estas proteínas en la región remanente de la Frontera F1 empleando oligonucleótidos específicos para cada mutante y lo comparamos contra el nivel de enriquecimiento observado en las células silvestres (Figura 17f). Al evaluar la frontera F1 de *Notch* empleando oligonucleótidos específicos para cada mutante, observamos que la disrupción parcial o total de la frontera F1 resulta en la pérdida de la marca de histonas H3K27ac en la región 5' de *Notch*. En particular, observamos que la eliminación de la Región Proximal al Promotor de *Notch* (mutante *5pN-Δ102*) resulta en la disminución de

la presencia de ARN Pol II, tanto pSer2 como pSer5, y de CTCF. La disminución de ARN Pol II es esperada dado que la mutante 5pN- Δ 102 pierde las primeras ~100 pb río arriba del inicio de la transcripción de *Notch*. Sin embargo, la disminución de la ocupación de CTCF en la mutante 5pN- Δ 102 es un resultado inesperado puesto que el motivo de unión al ADN para esta proteína no fue eliminado (Figura 16d), lo cual sugiere que secuencias de la Región Proximal al Promotor son importantes para el reclutamiento de CTCF. En el caso de la mutante para la Región Intergénica (mutante 5pN- Δ 183), se observa una disminución aún más pronunciada en el reclutamiento de ARN Pol II, tanto pSer2 como pSer5, además de la pérdida de unión de dCTCF, consistente con la remoción de su sitio de unión al ADN en esta mutante (Figura 16d y 17f). Finalmente, observamos que la eliminación de ambos segmentos en la mutante 5pN- Δ 343 resulta en la pérdida de unión para todas las proteínas evaluadas lo cual correlaciona con la fusión de dominios adyacentes y por lo tanto la inactivación de la frontera F1 (Figura 17f).

En mamíferos se ha descrito una interdependencia en el reclutamiento de Proteínas Arquitectónicas en las fronteras de Dominios Topológicos de tal forma que la disrupción de una frontera de dominio puede afectar el reclutamiento de Proteínas Arquitectónicas en una frontera adyacente⁷⁶. Para evaluar el efecto de la disrupción de la frontera F1 en la estructura de la cromatina de la frontera F2 y el Dominio 2 de *Notch* evaluamos el nivel de enriquecimiento para ARN Pol II y la marca de histonas H3K27ac en la frontera F2 así como al interior del Dominio D2 en todas las mutantes de la frontera F1 (Figura 17f). No observamos cambios en el nivel de enriquecimiento para ninguna de las proteínas evaluadas en la frontera F2 y al interior del Dominio 2. Estos resultados correlacionan con la ausencia de cambios topológicos para el Dominio 2 en las mutantes de la frontera F1 de *Notch* y sugieren que la frontera F1 no influencia la topología ni la estructura de la cromatina en la frontera F2 y el Dominio 2.

La pérdida de ARN Pol II que observamos en la mutante de la Región Intergénica (mutante 5pN- Δ 183) es incluso mayor que la observada al eliminar la Región Proximal al promotor y sugiere que esa secuencia es importante para el reclutamiento de la ARN Pol II (Figura 16d, 17f). Puesto que CTCF es el único motivo de unión a Proteínas Arquitectónicas identificado en la Región Intergénica, hipotetizamos que podría ser importante para el reclutamiento de ARN Pol II. Para evaluar la contribución de CTCF en el reclutamiento de proteínas nucleares se realizó un Ensayo de Cambio en Movilidad Electroforética (Electrophoretic Mobility Shift Assay; EMSA) empleando extractos proteicos nucleares de células S2R+. Se encontró que la mutación del sitio de unión a CTCF en la Región Intergénica de *Notch* resulta en una marcada pérdida de retardo de la banda mutante (Figura Suplementaria 5). Esto sugiere que el motivo de unión a CTCF afecta el reclutamiento de proteínas nucleares como la ARN Pol II (Vea Discusión).

En conjunto, nuestros datos demuestran que la inactivación de una frontera de dominios resulta en la fusión de dominios topológicos. La frontera F1 de *Notch* es un aislador cromatínico con módulos que de forma independiente afectan la

conformación 3D de sus dominios adyacentes. De forma relevante, la disrupción de la frontera y los efectos topológicos observados, correlacionan con la pérdida del reclutamiento de CTCF, ARN Pol II y H3K27ac.

VI. La disrupción de la frontera F1 de *Notch* afecta la transcripción génica

Para evaluar los efectos transcripcionales asociados a la inactivación de la frontera F1 de *Notch*, realizamos experimentos de RT-PCR en tiempo real empleando ARN de células silvestres y todas las mutantes de la frontera F1 y pares de oligonucleótidos dirigidos contra los exones localizados en los Dominios 1 y 2 de *Notch* (Figura 18a,b).

Observamos una disminución en la transcripción de los exones del Dominio 1 para todas las mutantes de la frontera F1 de *Notch* (Figura 18b). La pérdida de transcripción dentro del Dominio 1 es consistente con la pérdida de unión de ARN Pol II en la región 5' intergénica de *Notch* y correlaciona con la pérdida de contactos dentro del Dominio 1 y la ganancia de contactos ectópicos entre el extremo 5' UTR de *Notch*, que contiene el sitio de inicio de la transcripción, con el dominio kirre-D2 (Figura 2d). En contraste, la transcripción en el Dominio 2 no se ve afectada o bien muestra una tendencia al incremento (mutante *5pN-Δ183*) (Figura 18b). En el caso de la mutante *5pN-Δ183* observamos que el incremento en la transcripción correlaciona con la ganancia de interacciones entre secuencias del Dominio 2 así como con un incremento modesto en los niveles de ARN Pol II pSer2 en el exón 6 (Figura 18d y 17f). Por lo tanto, la disrupción de la frontera F1 resulta en la disminución de la transcripción exclusivamente en el Dominio 1 mientras que el Dominio 2 se comporta como una unidad topológica y transcripcional independiente.

Las eliminaciones generadas en la frontera F1 también resultan en la ganancia de contactos ectópicos entre secuencias de Dominios Topológicos adyacentes a *Notch*, por lo cual evaluamos el nivel de transcripción para dos de las siete isoformas de *kirre* (*kirre-full* y *kirre-med*) cuyos promotores coinciden con las fronteras 5' de los Dominios Topológicos kirre-D1 y D2 respectivamente, así como del gen *rst*, que se localiza entre la frontera 3' del dominio kirre-D1 y la frontera 5' del dominio kirre-D2 (Figura 18a).

Observamos una disminución de ~50% en la transcripción de la isoforma *kirre-med* exclusivamente en la mutante *5pN-Δ343* mientras que el gen *rst* no presenta cambio en su estado transcripcional (Figura 18c). El efecto en la transcripción de *kirre-med* en la mutante *5pN-Δ343* puede estar relacionado con la fusión de los Dominios Topológicos kirre-D2 y el Dominio 1 de *Notch* a consecuencia de la inactivación de la frontera F1 (Figura 17e). En el caso del transcrito *kirre-full*, observamos una disminución de al menos el 50% en la mutante *5pN-Δ102* y hasta de un 75% en la mutante *5pN-Δ343* (Figura 18c). El cambio en la transcripción observado para *kirre-full* se acompaña de la ganancia de contactos ectópicos entre el promotor de *kirre-full* y regiones genómicas

distales río abajo y que presentan ocupación de ARN Pol II y del co-factor de Proteínas Arquitectónicas, CP190 (Figura 18e). Por ejemplo, en la mutante *5pN-Δ343* el promotor de *kirre-full* establece contactos ectópicos con la frontera F2 de *Notch* así como con una región río abajo a una distancia de ~1 Mb y que se encuentra ocupada por ARN Pol II y CP190 (Figura 18e).

Finalmente, evaluamos el estado transcripcional del gen *dnc*, localizado río abajo de *Notch* y observamos una disminución de ~50% en los niveles de transcrito en la mutante *5pN-Δ183*, que a su vez se acompañan de una ganancia de contactos ectópicos entre el promotor de *dnc* y el Dominio Topológico *kirre-D2* (Figura 18c,d). Por lo tanto, la inactivación de la frontera F1 de *Notch* se acompaña de cambios transcripcionales.

VII. La eliminación de la frontera F2 provoca la fusión de los dominios de *Notch*

El Dominio 2 de *Notch* no muestra cambios topológicos ni transcripcionales ante la pérdida de la frontera F1, por lo cual proponemos que es una unidad topológica y transcripcional independiente de la frontera F1 pero que depende de la frontera F2 de *Notch*. La región genómica en donde se localiza la frontera F2 corresponde a un elemento *enhancer* caracterizado por la presencia de ARN Pol II así como de marcas de histonas asociadas a elementos *enhancer*, como la H3K4me1 y la H3K27ac (Figura 16a y 19a). Para evaluar la contribución de la frontera F2 en la organización tridimensional del locus de *Notch*, eliminamos la región *enhancer* mediante CRISPR-Cas9 y evaluamos el efecto topológico por *in nucleus* Hi-C.

La eliminación de la frontera F2 resulta en la fusión de los Dominios Topológicos de *Notch* (Figura 19b). En particular observamos un incremento en los contactos entre los dominios de *Notch* (*p*-valor $2.2e-16$) acompañados por un decremento en los contactos al interior del Dominio 2 (*p*-valor 0.0003939) (Figura 19c). Además, análisis de 4C virtual usando como anclas (viewpoints) a regiones genómicas que corresponden a los Dominios 1 y 2 muestran un incremento en los contactos entre dominios, en particular, observamos una ganancia de contactos entre la región que corresponde a los exones 7-9 del Dominio 2 con todo el Dominio D, hasta la frontera F1 (Figura 19d). La identificación computacional de fronteras a nivel de genoma completo confirma la pérdida de la frontera F2 pero no de la frontera F1, y la re-organización del locus de *Notch* en un solo dominio topológico (Figura 19b). Estos datos sugieren que el *enhancer* del intrón 5 y que es detectado como la frontera F2 de *Notch* por experimentos de Hi-C es fundamental para la organización tridimensional del locus.

VIII. La frontera F2 *Notch* es necesaria para el reclutamiento de Proteínas Arquitectónicas y ARN Pol II y afecta la transcripción de todo el locus de *Notch*

Similar a lo observado en las mutantes de la frontera F1, la eliminación del frontera F2 resulta en una disminución de la marca de histonas H3K27ac y de ARN Pol II en la zona que flanquea la región eliminada, consistente con la eliminación de un elemento enhancer (Figura 19e). De forma inesperada, observamos la pérdida significativa del enriquecimiento de ARN Pol II y H3K27ac en el exón 6, lo cual sugiere que el posicionamiento de ARN Pol II y el enriquecimiento de marcas de histonas en el exón 6 es dependiente del enhancer en el intrón 5 (Figura 19e). Además, observamos una tendencia a la baja de ARN Pol II y H3K27ac en la frontera F1 pero no de CTCF, lo cual sugiere que, a diferencia de la frontera F1, el elemento enhancer identificado como la frontera F2 es crucial para el reclutamiento de ARN Pol II y H3K27ac en elementos regulatorios del locus de *Notch* pero no así de CTCF.

La disminución de ARN Pol II en el locus de *Notch* a consecuencia de la eliminación de la frontera F2 sugiere efectos en la transcripción de *Notch*. En contraste con lo observado en las mutantes para la frontera F1, la inactivación de la frontera F2 de *Notch* resulta en un decremento de la transcripción a lo largo del locus (Figura 19f). En particular, la transcripción de los exones en el Dominio 2 es severamente afectada, ya que muestra un decremento de hasta 75% en comparación con las células silvestres, en contraste con lo observado en las mutantes para la frontera F1 (Figura 18b). La disminución de la transcripción en todo el locus así como la pérdida de ARN Pol II tanto en la región 5' intergénica y en el Dominio 2 son consistentes con la pérdida de un elemento enhancer en la frontera F2 que es fundamental para la transcripción y la organización tridimensional del locus.

Finalmente, observamos que la transcripción de los genes *kirre* y *dnc* también es afectada, y esto correlaciona con la ganancia de contactos ectópicos con regiones enriquecidas para ARN Pol II y CP190 (Figura 19f,g). Por ejemplo, en las células mutantes para la frontera F2, el promotor de *kirre-full* gana interacciones de larga distancia con una región río abajo y ocupada por RNA Pol II y CP190 (Figura 19f).

Por lo tanto, nuestros datos sugieren que la pérdida de la frontera F2 de *Notch* resulta en la reconfiguración de la organización tridimensional del locus que se acompaña por cambios transcripcionales locales y de larga distancia. Además, observamos que la frontera F2 es importante para el reclutamiento de ARN Pol II al locus.

IX. La frontera F2 de *Notch* es necesaria para la formación de un Dominio Topológico de 1 Mb y afecta la regulación de los genes al interior

Nuestros datos de *in nucleus* Hi-C a 20 kb de resolución provenientes de células S2R+ silvestres muestran la existencia de un Dominio Topológico de 1Mb cuyas fronteras coinciden con el locus de *Notch* y un desierto génico de ~100 kb. Este Dominio Topológico que denominamos Mega-dominio presenta diversas particularidades:

1. Las fronteras del Mega-dominio se contactan en el espacio nuclear. Esto se puede observar como una franja o “rectángulo” en la parte superior del Mega-dominio en un mapa de calor de datos de Hi-C a 20 kb de resolución (Figura 20a).
2. Las fronteras del Mega-dominio forman parte del Compartimento B (Figura 20a).
3. El locus de *Notch* contacta al menos dos segmentos del Mega-dominio. Esto se puede observar como una franja de mayor señal que parte de *Notch* y se extiende a lo largo del extremo izquierdo del Mega-dominio (Figura 20a).
4. Las secuencias al interior del Mega-dominio se organizan en Dominios Topológicos de menor tamaño y establecen contactos de larga distancia entre ellos (Figura 20a).
5. La frontera F2 de *Notch* y que se encuentra localizada en el enhancer del intron 5 contacta significativamente diversas regiones al interior y a los costados del desierto génico pero no regiones al interior del Mega-dominio (Figura 20b).

En datos de Hi-C provenientes de la mutante para la frontera F2 de *Notch* observamos la pérdida del Mega-dominio, lo cual sugiere que la frontera F2 es suficiente y necesaria para mediar la formación de dicha estructura. En particular, observamos la pérdida de contactos entre la frontera F2 y el desierto génico (Figura 20a).

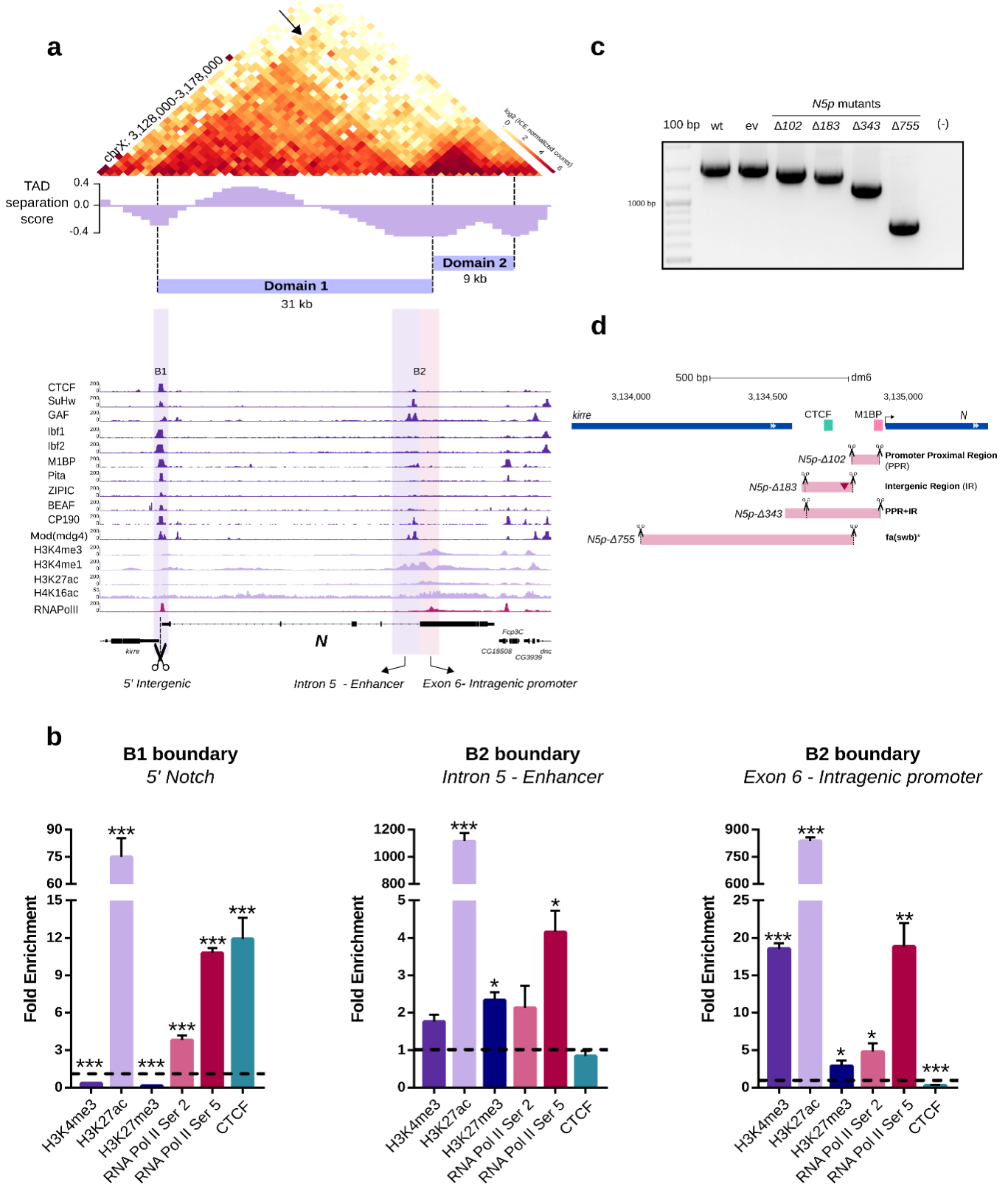
También observamos que los Dominios Topológicos al interior del Mega-dominio no presentan diferencias significativas, lo cual sugiere que su formación es independiente de la existencia del Mega-dominio, sin embargo, sí observamos la pérdida de contactos de larga distancia entre los Dominios Topológicos al interior del Mega-dominio (Figura 20a). Además, la pérdida del Mega-dominio se acompaña por un incremento de contactos ectópicos de larga distancia entre las secuencias contenidas al interior del Megadominio y las secuencias adyacentes como ocurre entre el locus *dnc* (originalmente contenido al interior del Mega-dominio) y el locus *kirre* río arriba de *Notch* (Figura Suplementaria 6). Finalmente observamos que la pérdida del Mega-dominio es específica de la mutante de la frontera F2 ya que dicha estructura se observa en cada una de las mutantes para la frontera F1 (Figura Suplementaria 6).

El Mega-dominio de *Notch* contiene ~70 genes. Sabemos que la inactivación de la frontera F2 afecta la organización tridimensional de *Notch* y su transcripción, en parte al afectar el reclutamiento de ARN Pol II a locus. Puesto que la frontera F2 es un enhancer, y dado que su eliminación provoca la pérdida del Mega-dominio

hipotetizamos que las células mutantes para la frontera F2 podrían presentar cambios en la transcripción de los genes contenidos al interior del Mega-dominio. Para evaluar los cambios transcripcionales a causa de la eliminación de la frontera F2 a nivel de genoma completo realizamos experimentos de RNA-seq empleando muestras de ARN total de células S2R+ silvestres y mutantes para la frontera F2 (Figura 20c). Encontramos que el 67% de los genes contenidos en el Mega-dominio (34 de 59) se encuentran expresados diferencialmente entre ambas condiciones (q -valor <0.05). De forma interesante observamos que la gran mayoría de los genes que cambian su expresión (31 de 34) lo hacen con una tendencia hacia la baja en la mutante de la frontera F2, sugiriendo que el enhancer en la frontera F2 de *Notch* podría estimular la transcripción de los genes al interior del Mega-dominio o bien la formación del Mega-dominio *per se* podría influir en el establecimiento de interacciones regulatorias adicionales entre promotores y enhancers al interior del Mega-dominio. Entre los genes que presentan una baja en su nivel de transcripción se encuentran factores transcripcionales como *Myc* y *Mnt*, así como *tlk* que codifica para una serina/treonina quinasa que interactúa con remodeladores de la cromatina y *VhaAC39-1* que promueve la señalización de *Notch* en discos imaginales. También validamos nuestro RNA-seq al realizar RT-PCR en tiempo real para un subconjunto de genes localizados en el Mega-dominio.

En conjunto, nuestros datos sugieren que la frontera F2 de *Notch* es un elemento genético con dos actividades topológicas: por una parte es importante para el establecimiento de la estructura de dos dominios de *Notch* al mismo tiempo que es necesario para la formación de un Mega-dominio de contacto a través de establecer contactos de larga distancia con un desierto génico de ~100 kb río debajo de *Notch*. De forma importante, la pérdida del Mega-dominio impacta la transcripción de los genes localizados al interior del dominio, sugiriendo que la topología del genoma puede afectar la expresión génica.

Figura 16. La organización tridimensional de *Notch* en células S2R+



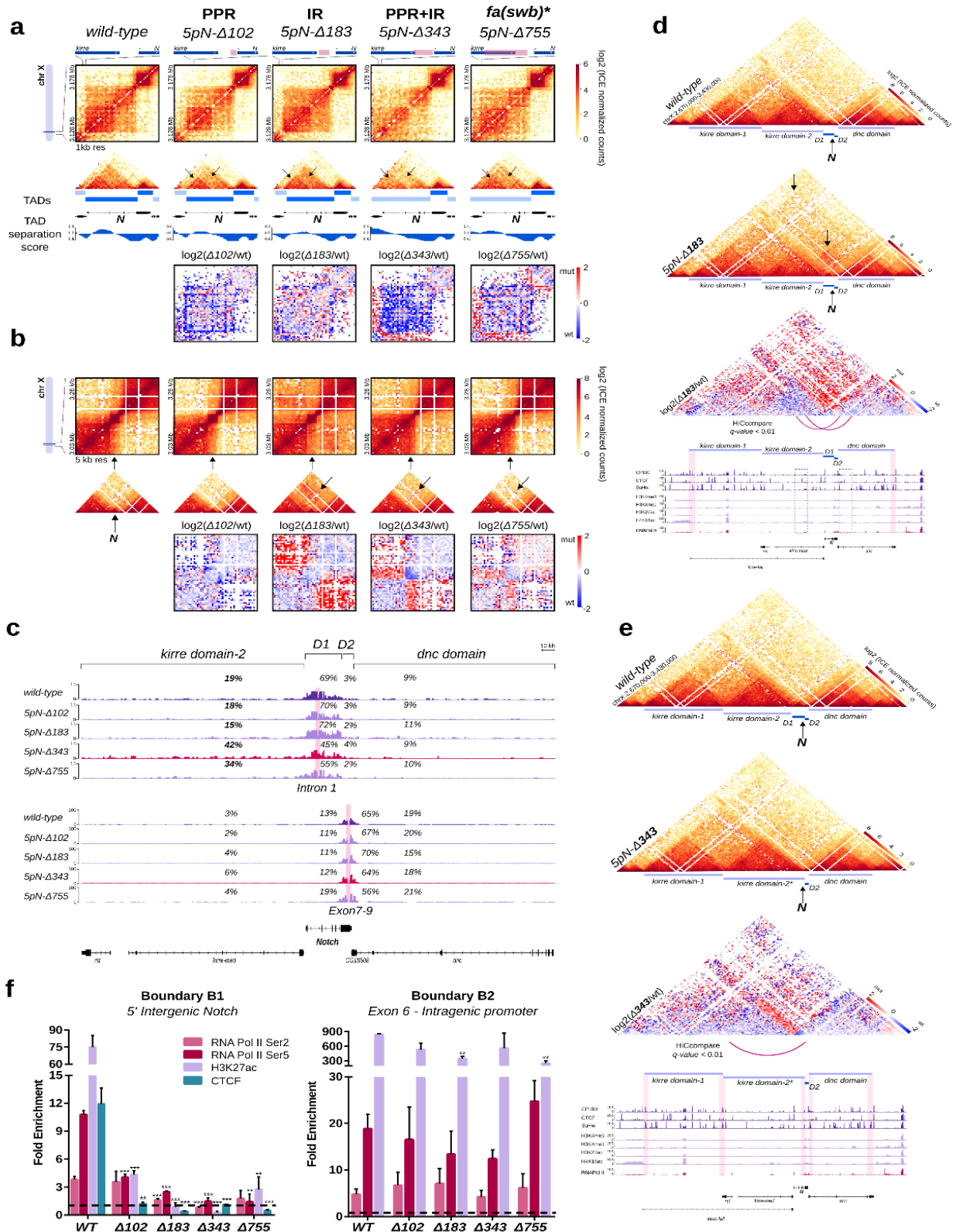
a Mapa de Hi-C a una resolución de 1 kb centrado en *Notch*. El locus de *Notch* se divide en dos dominios topológicos denominados como Dominio 1 y Dominio 2 presentados en la figura como Domain 1 y Domain 2 respectivamente. La frontera 5' del Dominio 1 (B1; Boundary 1 presentada en el texto como F1) interactúa con todo el locus de *Notch* (ver flecha). Debajo del mapa de Hi-C se muestran datos públicos de ChIP-seq para Proteínas Arquitectónicas (PAs), ARN Pol II y marcas de histonas para las células S2/S2R+. La posición de las fronteras de Dominios Topológicos en *Notch* está resaltada y se representan como B1 y B2. La frontera B2 es presentada en el texto como F2.

b Experimentos qChIP en células S2R + utilizando diferentes anticuerpos y oligonucleótidos para la región 5' intergénica (izquierda), el enhancer que sobrelapa con la frontera B2 (centro) y el exón 6 (derecha). Se presenta el enriquecimiento contra IgG. Las diferencias significativas con respecto a IgG se calcularon utilizando una prueba t. valor p * <0.05, ** <0.01, *** <0.001. Las barras de error representan el error estándar de la media (s.e.m.) de cuatro réplicas (n = 4).

c Genotipificación de clones mutantes CRISPR con eliminaciones que abarcan la frontera B1. Tamaño de amplicón esperado para el alelo silvestre (wt): 1500 pares de bases (pb). ev, vector vacío El número asociado con cada mutante CRISPR representa el número de pares de bases eliminados.

d Representación esquemática de clones mutantes CRISPR para la frontera B1. Los rectángulos rosados representan las regiones eliminadas. Las tijeras indican la posición de los sgRNAs utilizados para la edición del genoma mediada por CRISPR. El triángulo invertido rojo representa una inserción de 28 pb en la mutante 5pN-Δ183. fa (swb) * indica que la mutante 5pN-Δ755 es similar al alelo de fa (swb). Los sitios de unión de PAs detectados por FIMO (valor p <0.0001) se muestran como cuadros para CTCF y M1BP.

Figura 17. La eliminación de la región 5' intergénica de *Notch* da como resultado la fusión de Dominios Topológicos



a (Arriba) mapa de Hi-C a una resolución de 1 kb de una región de 50 kb centrada en *Notch*. (Centro) Representación triangular de mapas de Hi-C y TAD score para cada mutante. Las flechas indican regiones con diferencias en interacciones. (Abajo), mapas de Hi-C de las diferencias en log2 en la frecuencia de interacción entre los mutantes CRISPR y la condición silvestre (wild-type). Rojo indica ganancia de contactos en la mutante, y azul pérdida de contactos en la mutante. PPR, Región Proximal al Promotor; IR, Región intergénica. *fa* (swb) * indica que el mutante 5pN-Δ755 es similar al alelo *fa*(swb).

b (Arriba) mapas de Hi-C a una resolución de 5 kb de una región de 250 kb centrada en *Notch*. La flecha negra en la parte inferior de la gráfica indica la posición de *Notch*. (Centro) Representación triangular de mapas de Hi-C. Las flechas indican regiones con diferencias en las interacciones genómicas. (Abajo) mapas de Hi-C de las diferencias en log2 en la frecuencia de interacción entre los mutantes CRISPR y la condición silvestre (wild-type). Rojo indica ganancia de contactos en la mutante, y azul pérdida de contactos en la mutante.

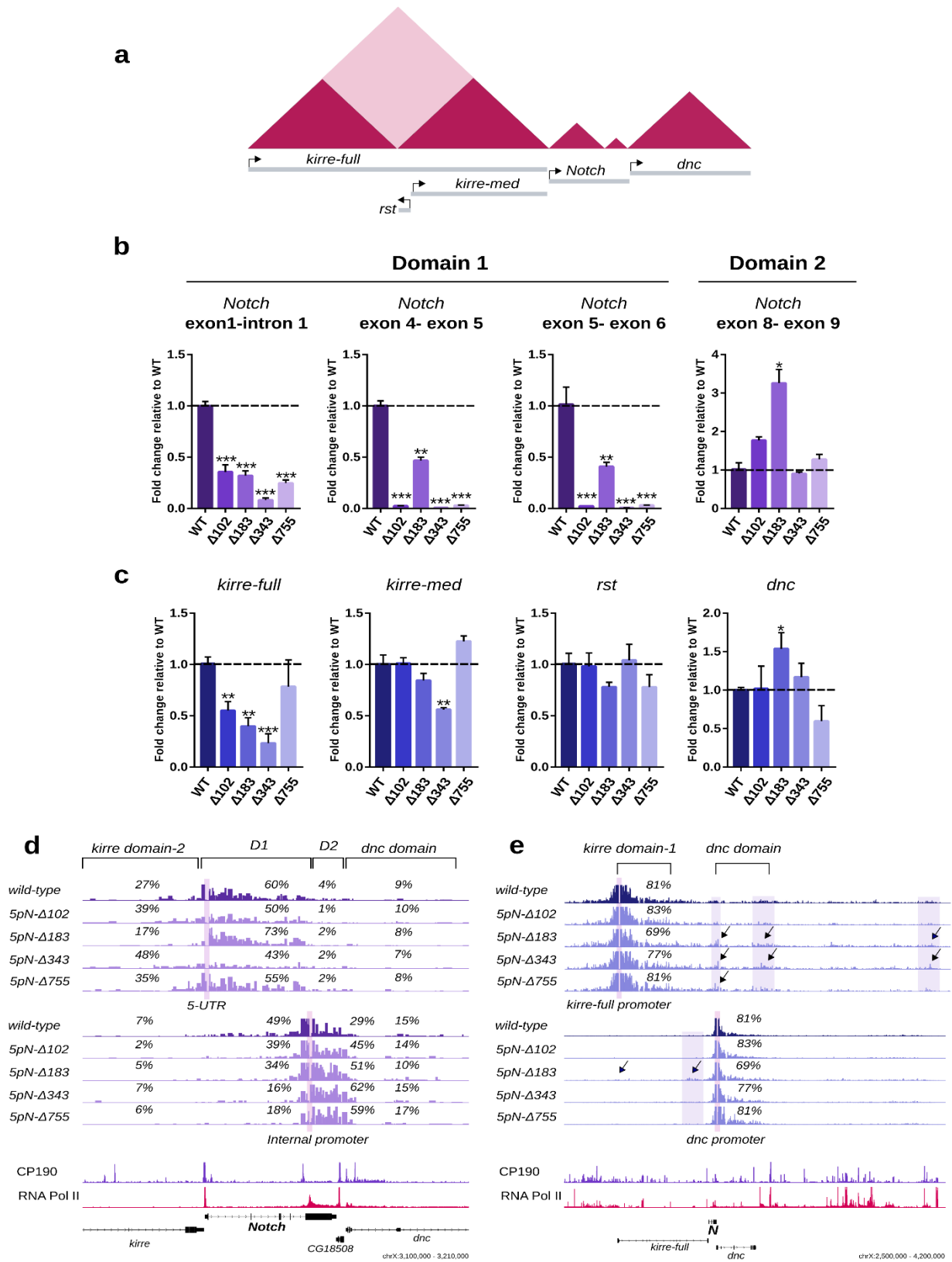
c Virtual-4C para los mutantes CRISPR y la condición silvestre utilizando *viewpoints* (rectángulos rosados) dentro del Dominio 1 y 2 de *Notch*. Se muestra el porcentaje de interacciones entre los *viewpoints* y el dominio *kirre-2*, los dominios de *Notch* y el dominio *dnc*.

d Mapas de Hi-C a una resolución de 5 kb para la condición silvestre (wild-type) y la mutante 5pN-Δ183 de una región de 760 kb. Las flechas indican regiones de ganancia de interacciones en el mutante. Se muestra un mapa de Hi-C de la diferencia log2 en la frecuencia de interacción entre la condición silvestre y la mutante 5pN-Δ183. También se muestran interacciones ectópicas detectadas en la mutante 5pN-Δ183 (arcos). Se muestran datos de ChIP-seq para Proteínas Arquitectónicas, marcas de histonas y ARN Pol II. Note que las interacciones ectópicas entre los loci *kirre* y *dnc* están restringidas por el límite 5' del dominio *kirre-1* y el límite 3' de *dnc* (resaltado en el track ChIP-seq).

e Mapas de Hi-C a una resolución de 5 kb para la condición silvestre (wild-type) y la mutante 5pN-Δ343 de una región de 760 kb. Se muestra un mapa de Hi-C de la diferencia log2 en la frecuencia de interacción entre la condición silvestre y la mutante 5pN-Δ343. Debajo se muestran interacciones ectópicas (arco) y tracks de ChIP-seq para Proteínas Arquitectónicas, marcas de histonas y ARN Pol II. Las regiones resaltadas corresponden a los límites del dominio. Note que la eliminación de la región 5' intergénica de *Notch* da como resultado la fusión del dominio *kirre-2* y el Dominio 1 de *Notch*.

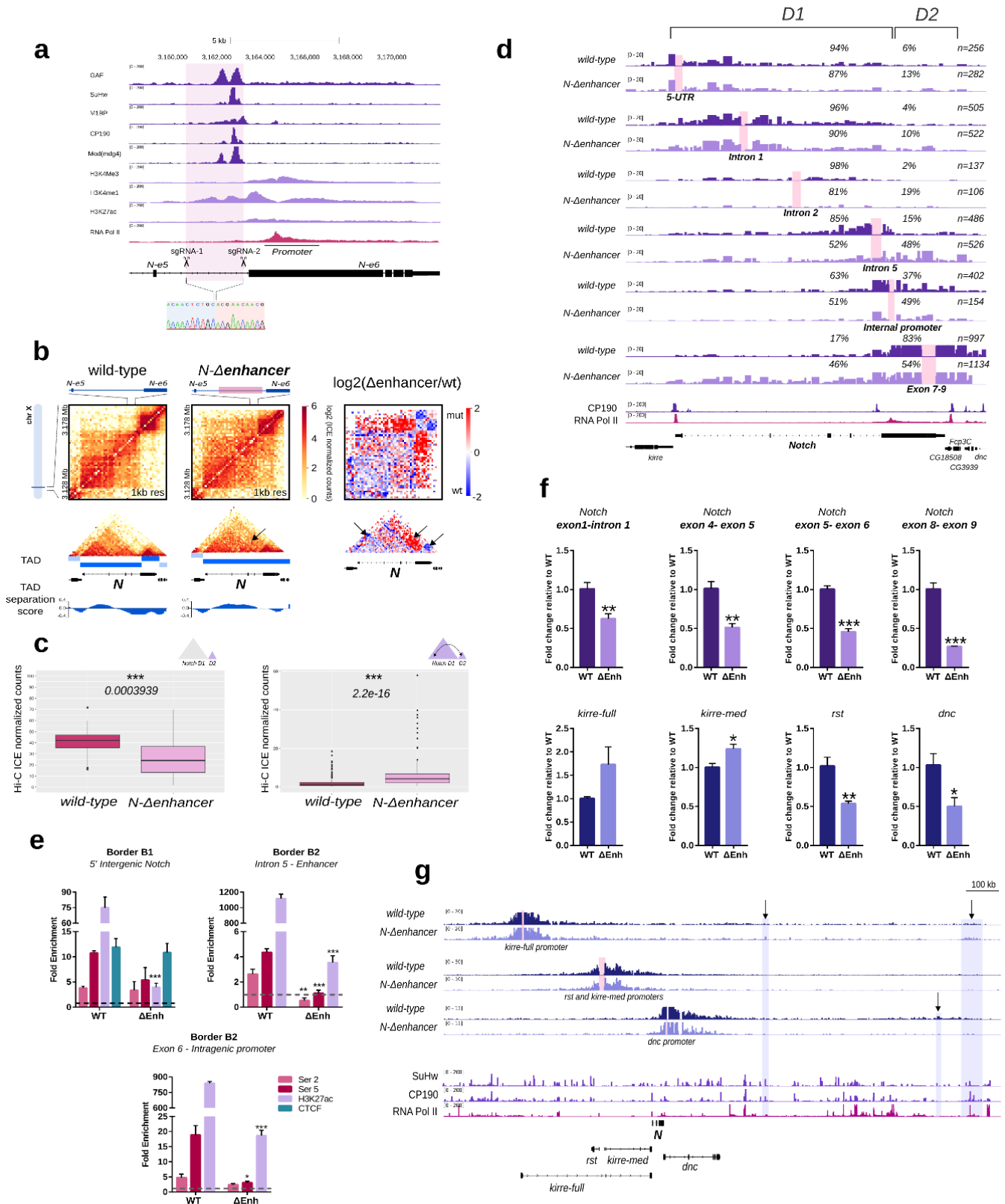
f Experimentos qChIP en células silvestres S2R + y clones mutantes CRISPR. Se presenta el enriquecimiento contra IgG. Las diferencias significativas entre mutantes y la silvestre se calcularon utilizando una prueba t. valor p * <0.05, ** <0.01, *** <0.001. Las barras de error representan el error estándar de la media (s.e.m.) de cuatro réplicas (n = 4).

Figura 18. La eliminación de la región 5' intergénica de *Notch* afecta la transcripción



- a** Representación esquemática los Dominios Topológicos adyacentes a *Notch* y los genes que analizamos por PCR cuantitativa en tiempo real (RT-qPCR).
- b** Transcripción de *Notch* medida por RT-qPCR. Cuantificaciones de transcripción en cada mutante CRISPR usando tres pares de oligonucleótidos que abarcan el Dominio 1 y un par para el Dominio 2. Las diferencias significativas entre la condición silvestre y las mutantes CRISPR se calcularon usando una prueba t. n = 3, valor p * <0.05, ** <0.01, *** <0.001. Las barras de error representan el error estándar de la media (s.e.m.) de cuatro réplicas (n = 4).
- c** Transcripción de genes ubicados en Dominios Topológicos adyacentes a *Notch*.
- d** Virtual-4C para la condición silvestre y mutantes CRISPR usando como viewpoints la región 5'UTR (arriba) y el exón 6 (abajo) de *Notch*. Se muestra el porcentaje de interacciones entre los viewpoints y las regiones dentro del dominio *kirre-2*, los dominios de *Notch* y el dominio *dnc*. También se muestran tracks de ChIP-seq para CP190 y RNA Pol II.
- e** Virtual-4C para la condición silvestre y mutantes CRISPR usando como viewpoints el promotor de la isoforma *kirre-full* (arriba) y el promotor de *dnc* (abajo). Se muestra el porcentaje de interacciones entre los viewpoints y los Dominios Topológicos en los que se encuentran. Las flechas indican regiones con interacciones ectópicas. También se muestran tracks de ChIP-seq para CP190 y RNA Pol II.

Figura 19. La eliminación la frontera intrónica resulta en la fusión de los dominios de Notch y defectos en la transcripción



a ChIP-seq para Proteínas Arquitectónicas, marcas de histonas y ARN Pol II para células S2/S2R + en el enhancer del intrón 5 de *Notch* que coincide con la frontera entre los Dominios Topológicos de *Notch*. Se resalta la región eliminada por CRISPR-Cas9. Debajo se muestra un electroferograma que representa el punto de corte de la eliminación.

b Mapas de Hi-C para la condición silvestre (izquierda) y la mutante (centro) a una resolución de 1 de una región de 50 kb centrada en *Notch*. También se muestra un mapa de Hi-C de las diferencias log2 en la frecuencia de interacción entre el tipo silvestre y el mutante (derecha). Las flechas indican regiones con cambios en las interacciones.

c Boxplots que representan el número de contactos detectados por Hi-C en la condición silvestre y al interior del Dominio 2 (izquierda) y entre dominios (derecha). Las diferencias significativas en el número de interacciones fueron determinadas con una prueba Wilcoxon. Valor $p^* < 0.05$, $** < 0.01$, $*** < 0.001$.

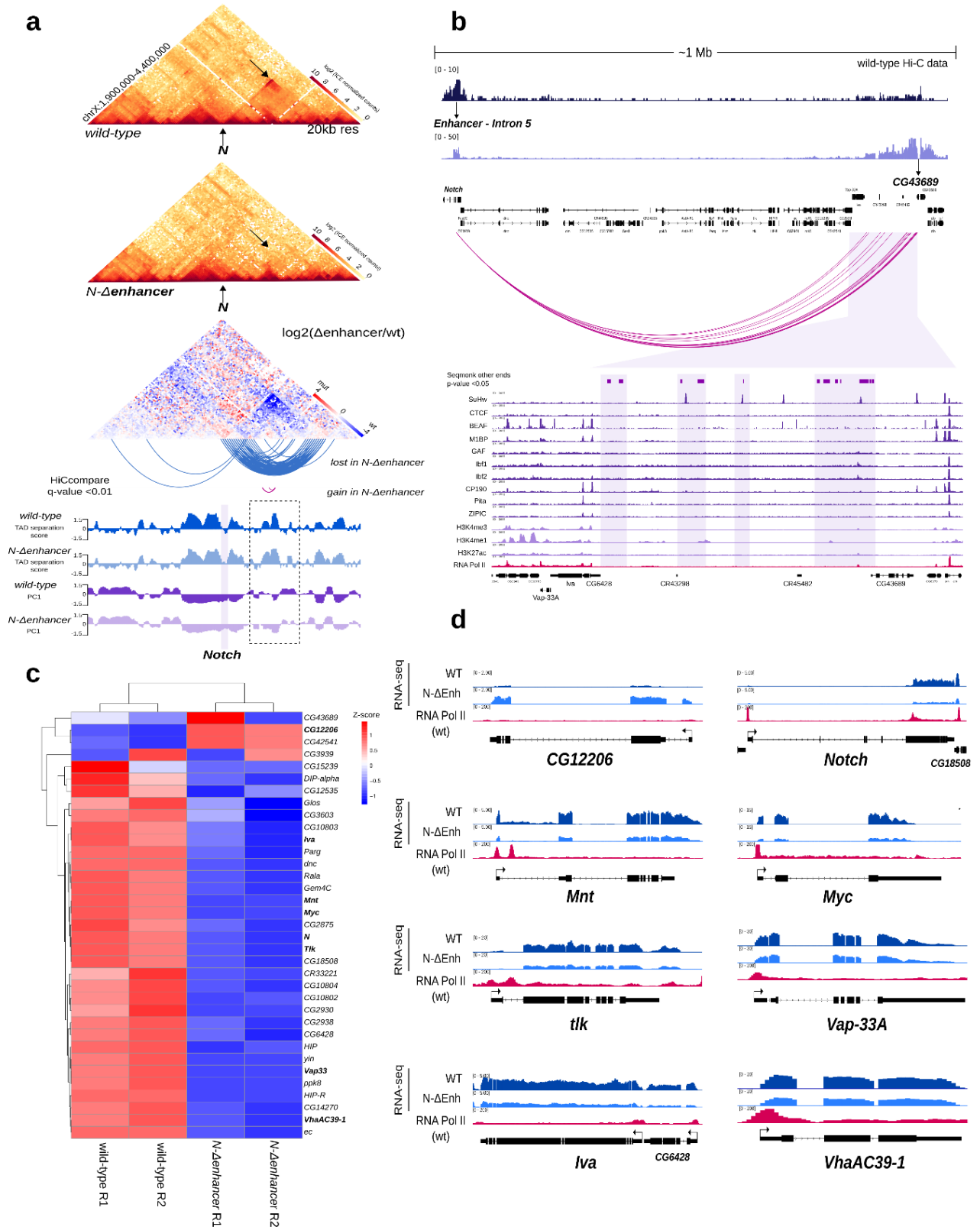
d Virtual-4C para la condición silvestre y mutante usando como viewpoints las secuencias que corresponden al 5' UTR, el Intron 1, Intron 2, Intron 5, el exón 6 y los exones 7-9 de *Notch*. Los porcentajes en cada track indican la proporción de interacciones entre el viewpoint y los dominios de *Notch*. Se muestran tracks ChIP-seq para CP190 y ARN Pol II.

e Experimentos qChIP en células de silvestres S2R + y mutantes. Se presenta el enriquecimiento contra IgG. Las diferencias significativas entre mutantes y la silvestre se calcularon utilizando una prueba t. valor $p^* < 0.05$, $** < 0.01$, $*** < 0.001$. Las barras de error representan el error estándar de la media (s.e.m.) de cuatro réplicas ($n = 4$).

f (Arriba), transcripción de *Notch* evaluada por RT-qPCR en la condición silvestre y mutante usando pares de oligonucleótidos que abarcan los Dominios de *Notch*. (Abajo), transcripción de genes ubicados en los Dominios Topológicos adyacentes a *Notch*. Para estimar diferencias significativas entre muestras, se utilizó una prueba t. Valor $p^* < 0.05$, $** < 0.01$, $*** < 0.001$. Las barras de error representan el error estándar de la media (s.e.m.) de cuatro réplicas ($n = 4$).

g Virtual-4C para la condición silvestre y mutante utilizando como viewpoints el promotor de la isoforma *kirre-full* (arriba), el promotor de *rst* y la isoforma *kirre-med* (medio) y el promotor de *dnc* (abajo). Las flechas indican regiones con interacciones ectópicas. Se muestran tracks ChIP-seq para SuHw, CP190 y ARN Pol II.

Figura 20. La frontera intrónica de *Notch* es necesaria para la formación del mega-dominio



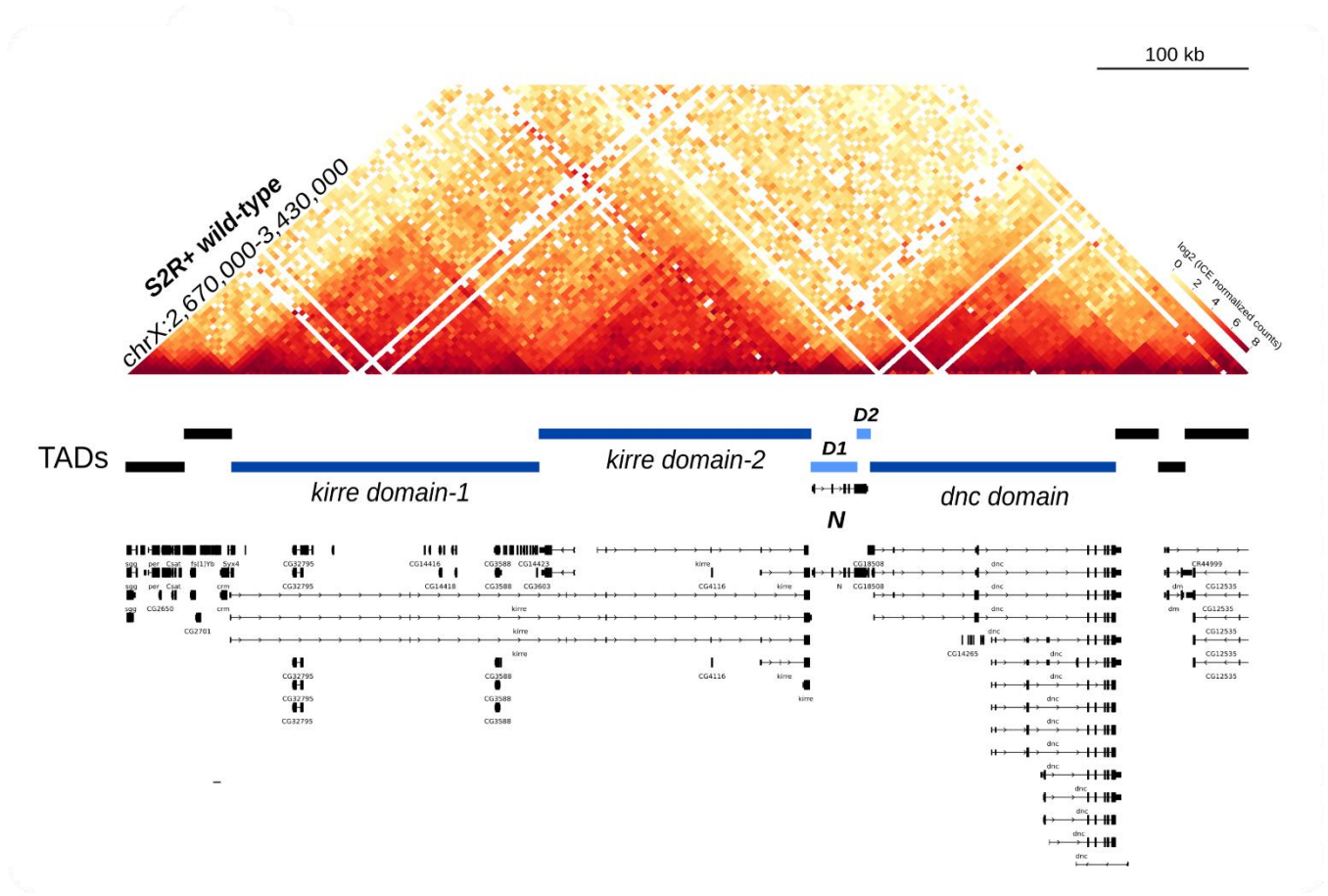
a Mapas de Hi-C a una resolución de 20 kb para la condición silvestre y la mutante. Se muestra un mapa de Hi-C de las diferencias \log_2 en la frecuencia de interacción entre el tipo silvestre y el mutante. HiCcompare detectó pérdida y ganancia de interacciones (arcos, valor $q < 0.01$). El cuadro punteado resalta una región con cambios en compartimentos.

b (Arriba), Virtual-4C para la condición silvestre usando como viewpoints la frontera intrónica de *Notch* (Enhancer-Intron 5) y el promotor del locus CG43689. (Centro) Regiones de interacción significativa con la frontera intrónica de *Notch* obtenidas a partir del análisis de datos Hi-C usando Seqmonk (valor $p < 0.01$). (Abajo), track de ChIP-seq para Proteínas Arquitectónicas, marcas de histonas y ARN Pol II. Las regiones resaltadas corresponden a fragmentos que interaccionan con la frontera intrónica de *Notch*.

c Mapa de calor que muestra los niveles de expresión detectados por ARN-seq para genes al interior del mega-dominio que cambian su nivel de expresión en la mutante. Note como la mayoría de los genes muestran una disminución en su nivel transcripcional.

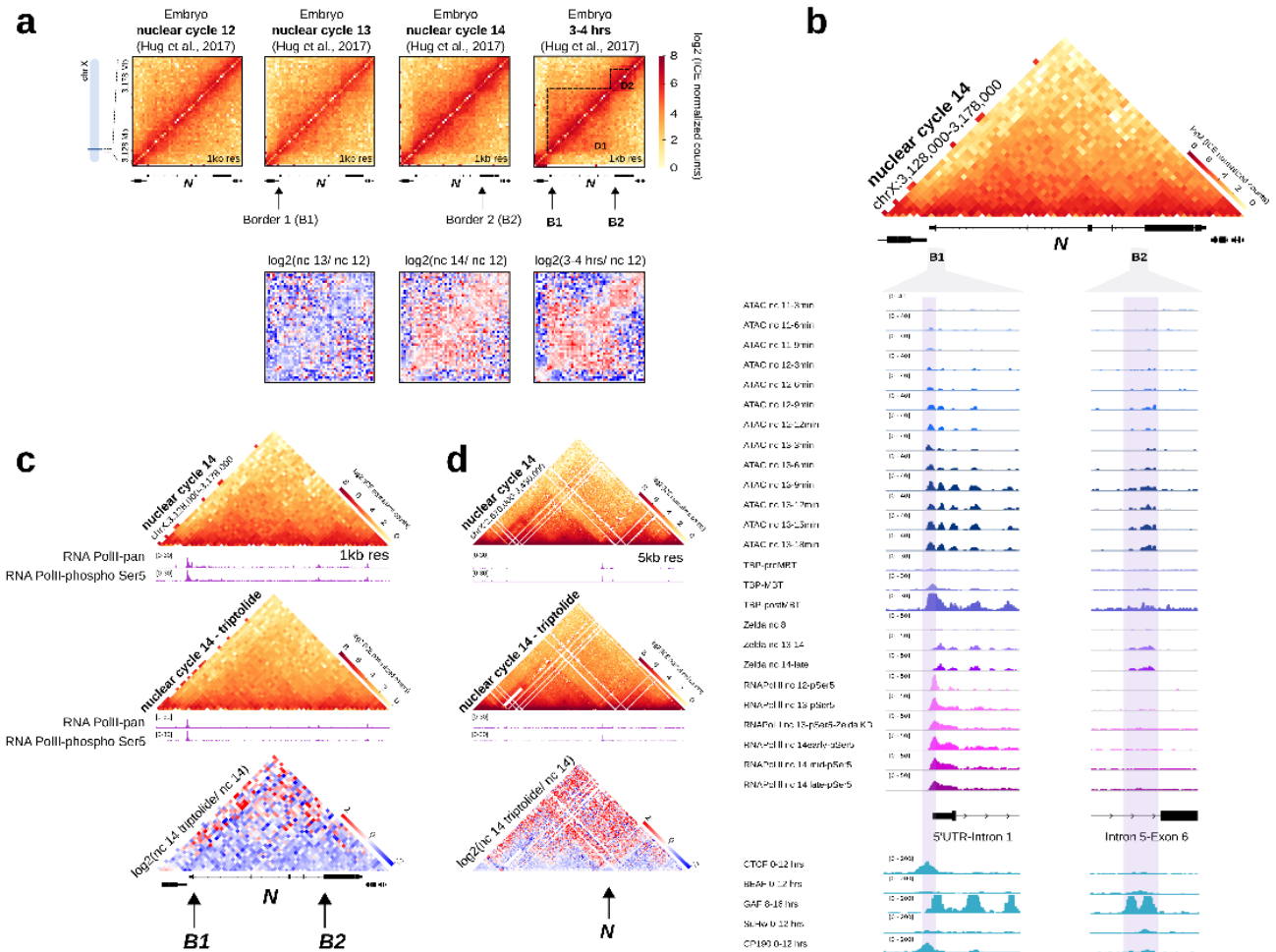
d Señal de ARN-seq para células silvestres y mutantes para un subconjunto de genes expresados diferencialmente en el mega-dominio. Se muestran tracks de ChIP-seq para ARN Pol II.

Figura Suplementaria 1. Paisaje topológico adyacente a Notch



Mapa de Hi-C a una resolución de 5 kb que muestra el paisaje topológico adyacente a *Notch*. Los Dominios Topológicos identificadas a una resolución de 1 kb se muestran debajo del mapa. El locus *kirre* se divide en dos Dominios Topológicos denominados dominio *kirre*-D1 y 2, mientras que el locus *dnc* está completamente contenido dentro de un Dominio Topológico.

Figura Suplementaria 2. La organización tridimensional de *Notch* se establece durante el desarrollo embrionario temprano



a Mapas de Hi-C a una resolución de 1 kb centrados en *Notch*. Hi-C de embriones en desarrollo (ciclo nuclear 12,13,14 y 3-4 horas de desarrollo; Hug et al., 2017). Las flechas negras indican la posición de las fronteras B1 y B2. Las líneas punteadas indican la posición de los Dominios D1 y D2 de *Notch*. Debajo se muestran los mapas de la diferencia en contactos de Hi-C entre el ciclo nuclear 12 (nc 12) y nc13, nc14 y 3-4 h.

b Representación triangular de un mapa de Hi-C de embriones en el ciclo nuclear 14 a una resolución de 1 kb centrados en *Notch*. Debajo del mapa de Hi-C se muestran datos públicos ATAC-seq y ChIP-seq para TBP, RNA Pol II y Zelda durante el desarrollo embrionario temprano (nc11-nc14) para las regiones identificadas como fronteras de dominios en locus *Notch*. También se muestran datos de ChIP-seq para Proteínas Arquitectónicas de embriones de 0 a 12 horas. MBT, transición a media blastula.

c Representación triangular de mapas de Hi-C a una resolución de 1 kb centrado en *Notch*. Arriba y en el medio, Hi-C del ciclo nuclear 14 embriones tratados con triptólidos o de tipo salvaje. Debajo de cada mapa de calor se muestran las track de RNA Pol II ChIP-seq de embriones tratados con triptólido y de tipo salvaje de Hug et al., 2017. Mapas de calor inferiores, Hi-C de la diferencia \log_2 en la frecuencia de

interacción entre el ciclo nuclear de tipo salvaje 14 y ciclo nuclear 14 embriones tratados con triptolida. Las flechas negras indican la posición de los límites B1 y B2 de Notch. Observe que la inhibición transcripcional global resulta en una disminución de las interacciones intradominio en el locus Notch; sin embargo, los TAD aún son visibles.

d Representación triangular de mapas de calor Hi-C a una resolución de 5 kb que cubre una región de 760 kb centrada en Notch. Arriba y en el medio, Hi-C del ciclo nuclear 14 embriones tratados con triptolida o de tipo salvaje. Debajo de cada mapa de calor se muestran las track de RNA Pol II ChIP-seq de embriones tratados con triptólido y de tipo salvaje de Hug et al., 2017. Mapas de calor inferiores, Hi-C de la diferencia log2 en la frecuencia de interacción entre el ciclo nuclear de tipo salvaje 14 y ciclo nuclear 14 embriones tratados con triptolida. La flecha negra indica la posición de Notch. Observe que la inhibición transcripcional global produce una disminución de las interacciones intradominio también en los TAD que flanquean el locus Notch; sin embargo, los TAD aún son visibles.

Figura Suplementaria 3. Diseño de eliminaciones mediante el sistema CRISPR-Cas9

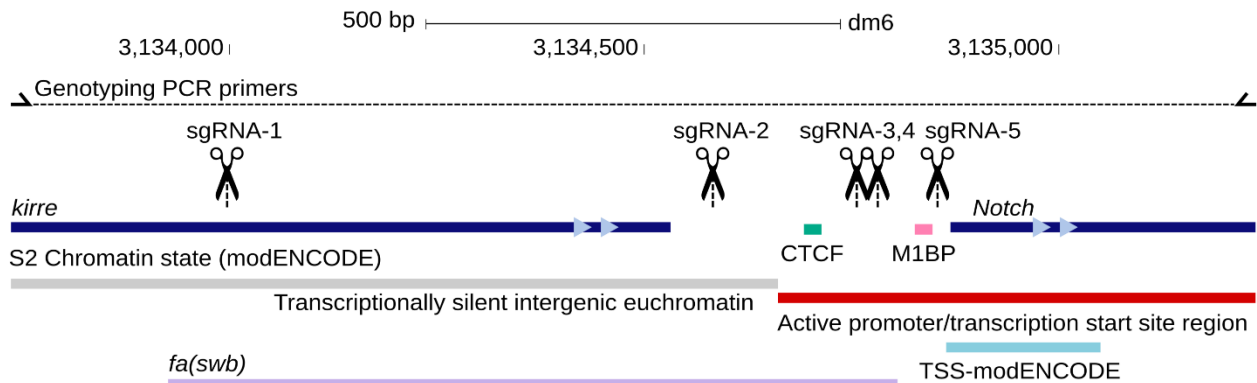
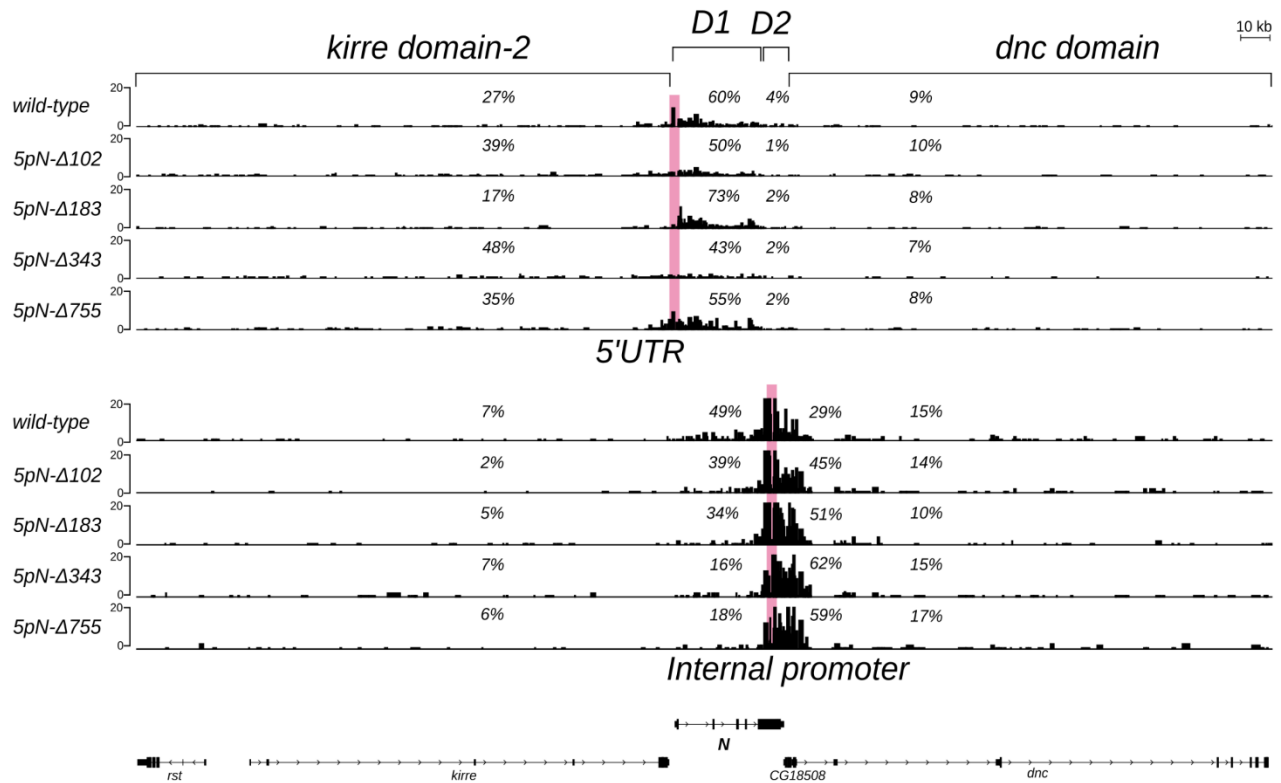
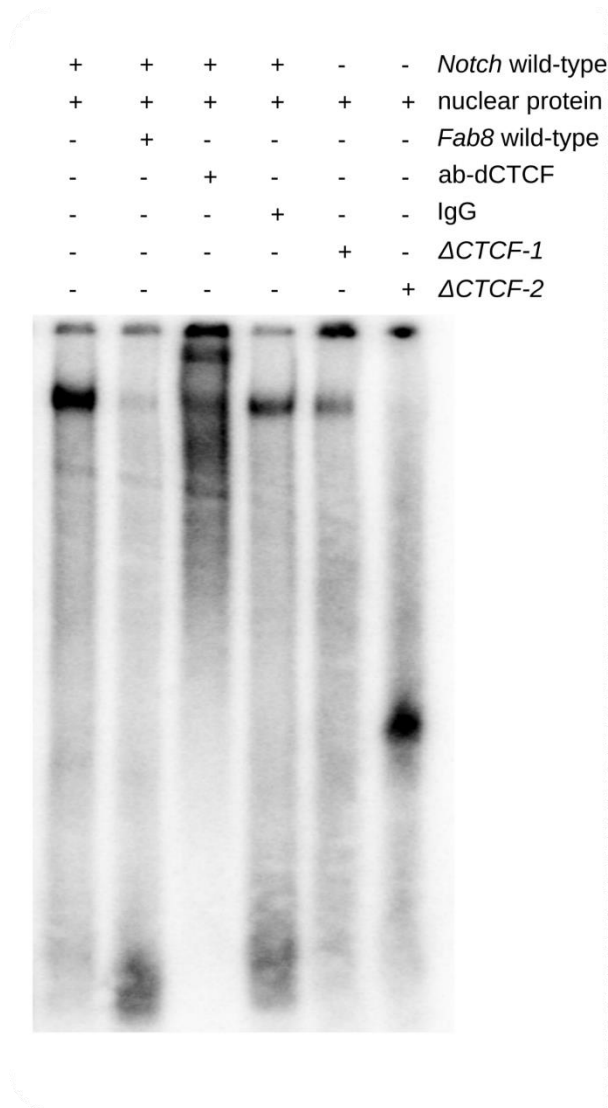


Figura Suplementaria 4. La disrupción de la frontera 5' de *Notch* resulta en la ganancia de contactos ectópicos



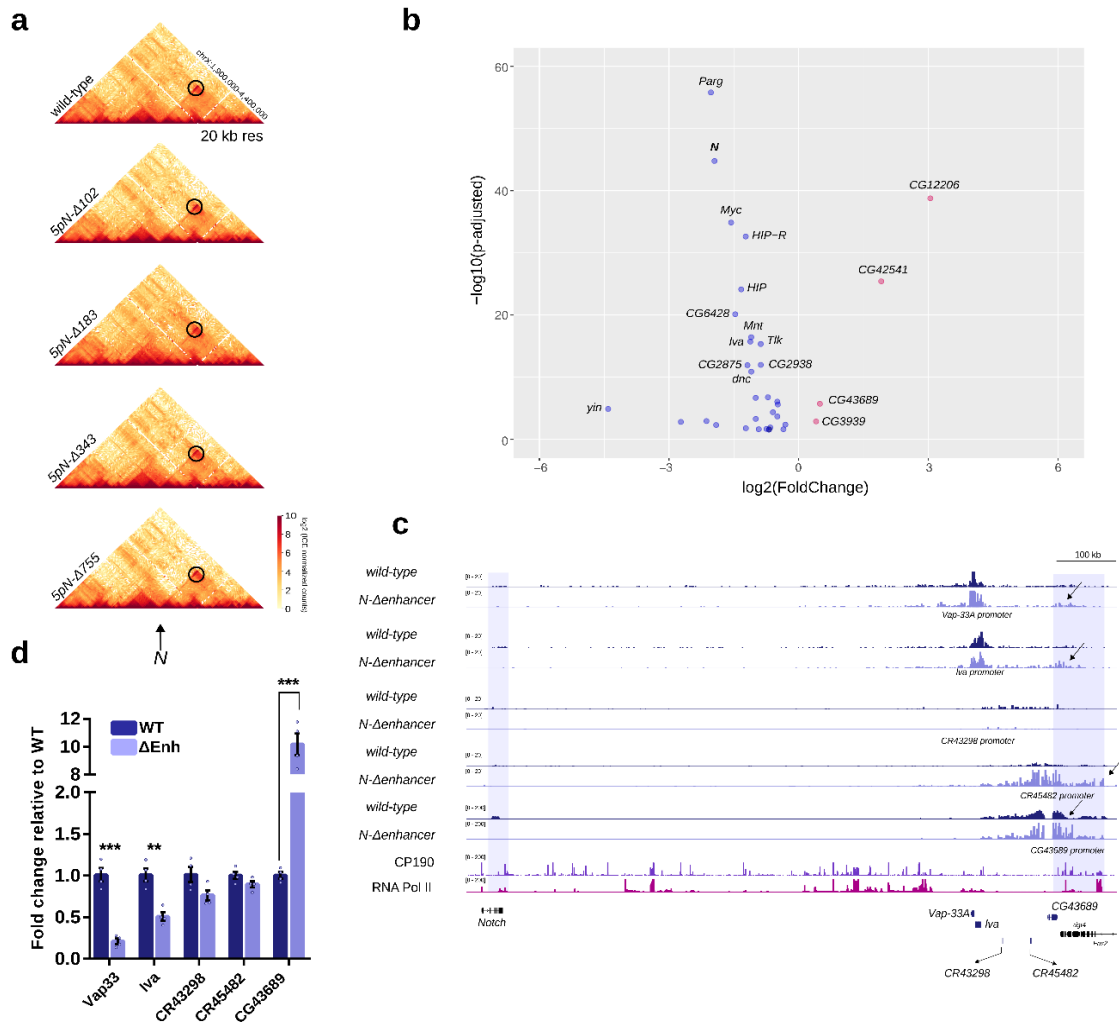
Virtual 4C de datos de Hi-C para la condición silvestre y mutantes CRISPR utilizando como viewpoints las secuencia del 5'UTR y el exón 6 de *Notch*. Los porcentajes en cada track indican la fracción de lecturas válidas entre el viewpoint y los Dominios Topológicos de *kirre*, *Notch* y *dnc*.

Figura Suplementaria 5. El motivo de unión a CTCF es necesario para el reclutamiento de proteínas nucleares



EMSA usando extractos nucleares de proteínas de la línea celular S2R + y oligonucleótidos marcados radioactivamente y presentados en la Tabla 2. Como control, se usó un oligonucleótido de 60 pb no marcado con un motivo de unión a CTCF validado presente en el insulador *Fab8*. De izquierda a derecha; 1, oligonucleótido con motivo silvestre de CTCF en *Notch*; 2, competencia del oligonucleótido con motivo silvestre de CTCF en *Notch* con el oligonucleótido *Fab8* no marcado; 3, ensayo de súper retardo con extractos nucleares de proteínas incubados con α -dCTCF y el con motivo silvestre de CTCF en *Notch*; 4, incubación de oligonucleótidos con motivo silvestre de CTCF en *Notch* con IgG; 5 y 6 oligonucleótidos mutantes CTCF incubados con extractos nucleares de proteínas. Observe que la mutación del sitio de unión a CTCF afecta el reclutamiento de proteínas nucleares.

Figura Suplementaria 6. Efectos topológicos y en transcripción asociados a la eliminación de la frontera F2 de *Notch*



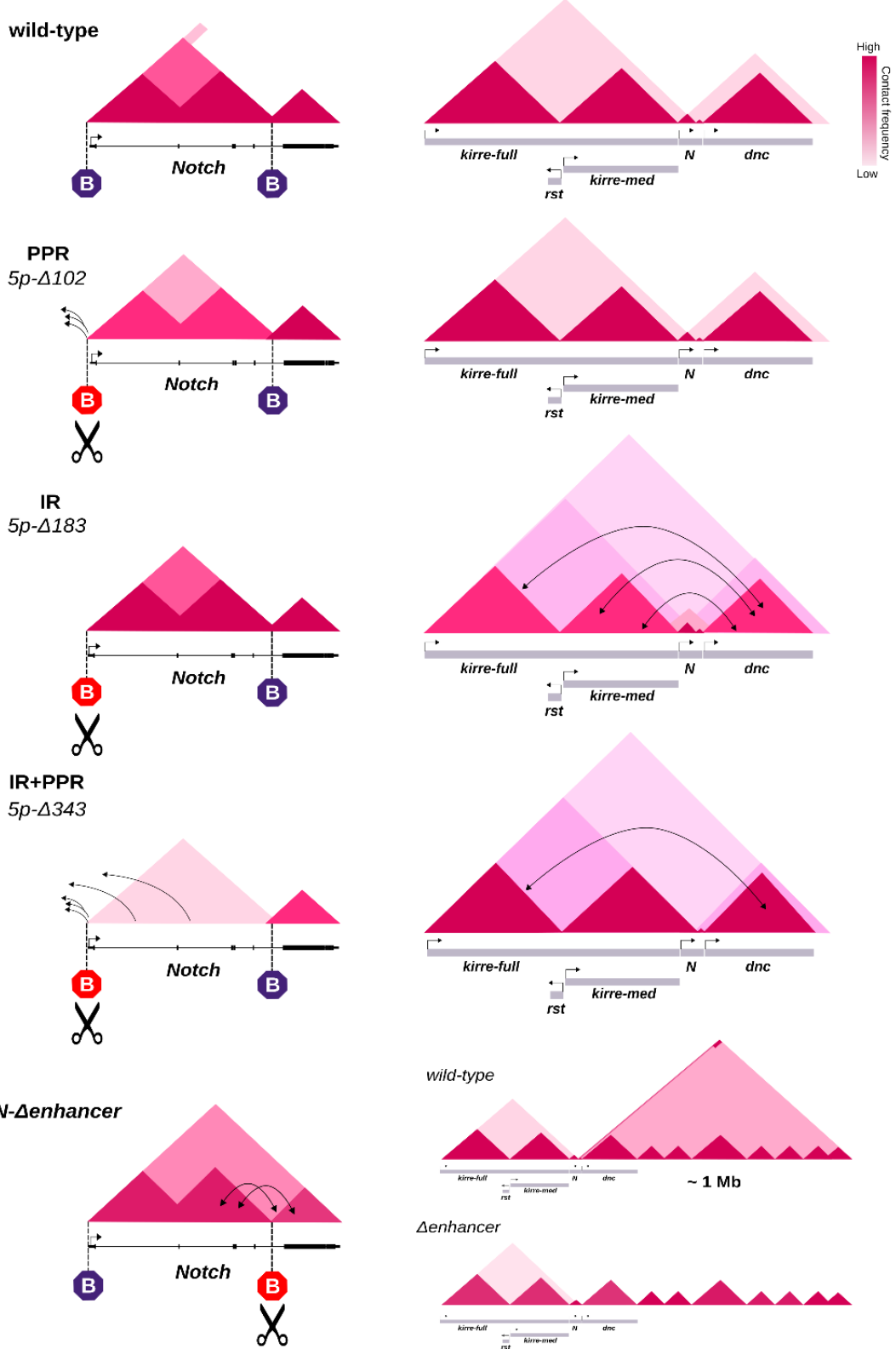
a Representación triangular de mapas de Hi-C a una resolución de 20 kb para una región de 2.5 Mb para células de tipo salvaje y los mutantes CRISPR de la frontera F1. Observe que mutaciones en la frontera F1 no afecta la organización del Mega-domino.

b Volcano-plot de los genes expresados diferencialmente dentro del Mega-domino. Los puntos azules representan genes regulados a la baja en la mutante de la frontera F2, mientras que los puntos rojos representan genes regulados positivamente en las células mutantes.

c Virtual 4C utilizando como viewpoints los promotores de genes ubicados al interior o cerca del desierto genético. Las flechas indican regiones con interacciones ectópicas. Se muestran tracks ChIP-seq para CP190 y ARN Pol II. Las regiones resaltadas corresponden a las fronteras del mega-domino.

d Transcripción de genes ubicados al interior o cerca del desierto genético. Las diferencias significativas entre la condición silvestre y la mutante se calcularon usando una prueba t. valor p * < 0.05 , ** < 0.01 , *** < 0.001 . Las barras de error representan el error estándar de la media de cuatro réplicas ($n = 4$).

Resumen gráfico de nuestros resultados



Capítulo 8.

Discusión y conclusiones

De inventarios a mecanismos

“In the long course of cell life on this earth it remained, for our age, for our generation, to receive the full ownership of our inheritance. We have entered the cell, the Mansion of our birth and started the inventory of our acquired wealth.”

Albert Claude
Nobel lecture
December 12, 1974

En este trabajo estudiamos las consecuencias topológicas y transcripcionales asociadas a la disrupción de fronteras de Dominios Topológicos en el locus de *Notch* en *Drosophila melanogaster*. Nuestros hallazgos sugieren que secuencias genéticas reconocidas por Proteínas Arquitectónicas y la ARNA Pol II funcionan como aisladores de la cromatina que organizan el genoma en Dominios Topológicos. La disrupción de fronteras de Dominios Topológicos afecta la topología del genoma, la transcripción y el reclutamiento de Proteínas Arquitectónicas así como de la ARN Pol II. Nuestros hallazgos tienen implicaciones importantes para nuestra comprensión de los mecanismos implicados en la organización tridimensional del genoma y el control de la expresión génica.

La organización tridimensional de *Notch* en interfase es consistente con el patrón de bandeo de cromosomas politénicos

Evidencia genética así como análisis de hibridación *in situ* de alta resolución del locus de *Notch* revelaron que la unidad transcripcional de *Notch* está contenida por completo en la banda 3C7 del cromosoma X⁶⁵⁻⁶⁷. En *Drosophila* la organización del genoma en Dominios Topológicos o TADs presenta una correspondencia notable con la organización en bandas e interbandas de cromosomas politénicos, sugiriendo que *Notch* está organizado en un Dominio Topológico. Nuestros datos de *in nucleus* Hi-C revelan que *Notch* está organizado en dos Dominios Topológicos de diferente tamaño en la línea celular embrionaria S2R+. Esta organización topológica tiene una correspondencia notable con la organización en dos sub-bandas de la banda 3C7 en cromosomas politénicos como lo revela un análisis de hibridación *in nucleus* de alta

resolución⁶⁷. También observamos que *Notch* se organiza en dominios de diferente tamaño lo cual es consistente con la distribución no simétrica del contenido de ADN dentro de la banda 3C7⁶⁷. Además, la inspección visual de los datos de los mapas de Hi-C a 1 kb de resolución revela niveles adicionales de organización al interior de los dominios de *Notch*.

Los Dominios Topológicos están conservados entre tipo celulares^{27,77} y emergen durante el desarrollo embrionario temprano tanto en *Drosophila* como en mamíferos^{62,69,78}. Consistente con estas observaciones, nuestros análisis de datos públicos de Hi-C para embriones de *Drosophila* revela la estructuración progresiva de *Notch* en dos Dominios Topológicos a lo largo del desarrollo embrionario temprano lo cual es consistente con nuestros datos en células S2R+ y con observaciones a nivel de genoma completo sobre la dinámica de organización tridimensional de los cromosomas⁶². Además, la organización de *Notch* en dos Dominios Topológicos es reportada también en una publicación independiente que hace uso de datos de ultra-resolución de *in nucleus* Hi-C derivados de embriones en *Drosophila*⁷⁹, sugiriendo que la organización tridimensional de *Notch* no varía entre tipos celulares y es potencialmente estable a lo largo del desarrollo. Finalmente nuestros datos en conjunto con observaciones derivadas de análisis genéticos y microscópicos sugieren fuertemente que la organización estructural observada en cromosomas politénicos es un reflejo de la organización del genoma en Dominios Topológicos.

Las fronteras de Dominios Topológicos con aisladores de la cromatina

¿Cuáles son los mecanismos que promueven la organización tridimensional del genoma en eucariontes y su impacto en la expresión génica? son preguntas de gran importancia en el campo de la biología de la cromatina y la regulación de la expresión génica. En particular, existe un intenso debate sobre la contribución que tiene la transcripción y los sitios de unión a Proteínas Arquitectónicas en la organización tridimensional del genoma. El análisis estructural por *in nucleus* Hi-C de nuestra colección de mutantes CRISPR con disrupción parcial o total de las fronteras de Dominios Topológicos en el locus de *Notch* proveen de evidencia experimental directa que apoya la existencia de elementos genéticos discretos y autónomos que son reconocidos por Proteínas Arquitectónicas y la ARN Pol II y que actúan como aisladores de la cromatina, siendo fundamentales para la organización de los cromosomas en Dominios Topológicos.

En particular, encontramos que una secuencia de 300 pb que comprende a la región intergénica entre los loci *kirre* y *Notch* es un aislador de la cromatina modular que funciona como frontera entre Dominios Topológicos. Descubrimos que segmentos adyacentes y no sobrelapantes de dicha secuencia funcionan como módulos independientes de aislamiento que activamente restringen las interacciones físicas entre los loci *kirre* y *Notch*, siendo necesaria la eliminación de todos los módulos para

observar la fusión de Dominios Topológicos. De forma importante, logramos correlacionar los efectos topológicos con la presencia de sitios de unión a Proteínas Arquitectónicas. Los efectos topológicos observados al eliminar la frontera F1 de *Notch* son consistentes con datos citológicos derivados de la mutante *fa(swb)* en donde una eliminación de ~900 pb que comprende parte de la región 5' intergénica de *Notch* resulta en la pérdida de una interbanda y la fusión de la banda 3C7 que contiene a *Notch* con la banda río arriba. Además, experimentos independientes como ensayos de bloqueo de *enhancer* y protección contra el silenciamiento epigenético⁷⁴ así como la observación de que la inserción ectópica de la región 5' intergénica de *Notch* es suficiente y necesaria para dividir una banda de cromosomas politénicos y crear una interbanda⁸⁰ apoyan la función de aislador de la frontera F1 de *Notch*. Por lo tanto, evidencia derivada de estudios citogenéticos, ensayos con genes reporteros y nuestros datos de *in nucleus* Hi-C para mutantes CRISPR demuestran que la región 5' intergénica de *Notch* es una frontera esencial para la organización tridimensional de *Notch*.

En el caso de la frontera F2, encontramos que esta corresponde a un elemento *enhancer*. Esta secuencia de 2.5 kb influye en la topología del genoma ya que su eliminación resulta en la fusión de los Dominios Topológicos de *Notch* así como en la pérdida de un Dominio Topológico de 1 Mb localizado río abajo de *Notch*.

Por lo tanto, nuestros datos de *in nucleus* Hi-C para mutantes de fronteras de *Notch* demuestran de forma conclusiva que las fronteras de Dominios Topológicos son esenciales para la organización de *Notch* y sugieren que regiones similares son fundamentales para la organización de los cromosomas en Dominios Topológicos en *Drosophila melanogaster*.

La unión de la ARN Pol II en las fronteras de Dominios Topológicos, pero no la transcripción, es importante para la actividad de aislamiento y para la organización en Dominios Topológicos

Reportes recientes sugieren un papel prominente para la transcripción y la ARN Pol II en la organización del genoma en Dominios Topológicos^{25,61,77}. Nuestros datos proporcionan observaciones importantes que apoyan el papel de la ARN Pol II en la actividad de fronteras y por lo tanto en la organización del genoma en Dominios Topológicos.

Primeramente, el re-análisis de datos públicos de Hi-C de embriones de *Drosophila melanogaster* en etapas tempranas del desarrollo sugiere que la frontera localizada en la región 5' de *Notch* (F1) se establece antes de la Activación del Genoma del Cigoto (ciclo nuclear 13) y por lo tanto es previo a la transcripción del locus. Sin embargo, se observa que la frontera localizada en el intrón 5 de *Notch* (F2) aparece durante Activación del Genoma del Cigoto (ciclo nuclear 14). La aparición de fronteras de

Dominios Topológicos en *Notch* correlaciona fuertemente con la adquisición de accesibilidad a la cromatina en etapas iniciales del desarrollo embrionario (ciclo nuclear 11) así como con la unión de proteínas como la ARN Pol II (ciclo nuclear 12), el factor de transcripción general TBP y el factor pionero Zelda. Puesto que no existen datos sobre la unión de otras Proteínas Arquitectónicas como CTCF en etapas tempranas del desarrollo embrionario, queda aún por determinar si estas se asocian a fronteras de dominios tempranos en el desarrollo embrionario. También observamos que la fuerza de las fronteras en *Notch* correlaciona con la accesibilidad de la cromatina y con la presencia de ARN Pol II lo cual es consistente con observaciones previas^{37,62}.

En segundo lugar, el re-análisis de datos públicos de Hi-C de embriones de *Drosophila melanogaster* en los cuales se inhibió la transcripción revelan que el bloqueo de la transcripción previa a la Activación Genómica del Cigoto resulta en un decremento en los contactos físicos al interior de los dominios de *Notch*, sin embargo las fronteras y los Dominios Topológicos de *Notch* aún son detectados, lo cual correlaciona con la retención de ARN Pol II en las fronteras de dominios sugiriendo que la ARN Pol II y otras proteínas son esenciales para la formación de Dominios Topológicos lo cual es consistente con observaciones previas derivadas de análisis de genoma completo en donde se correlaciona la ocupación de la ARN Pol II con la presencia de Dominios Topológicos⁶¹. Nuestros datos de *in nucleus* Hi-C para mutantes CRISPR de las fronteras de *Notch* también apoyan estas observaciones. La eliminación de la Región Proximal al Promotor (*5pN-Δ102*) resulta en una disminución de la transcripción de hasta un 80% con respecto a la condición silvestre pero el efecto topológico de esta mutante es discreto y solo observamos una disminución de las interacciones al interior del Dominio 1, lo cual sugiere que la transcripción podría influir en la compactación pero no en el establecimiento del dominio. También observamos que la eliminación de la Región Proximal al Promotor resulta en la reducción pero no la pérdida de ARN Pol II, lo cual sugiere que la ARN Pol II que permanece asociada a la región 5' intergénica de *Notch* podría ser suficiente para sustentar la actividad de frontera de dicha región. En soporte a esta propuesta, la fusión del Dominio D1 con el Dominio Topológico de *kirre*, localizado río arriba, correlaciona con la pérdida de ARN Pol II de la región 5' intergénica de *Notch*. Observamos una tendencia similar cuando eliminamos la frontera F2 de *Notch*, ya que la fusión de los Dominios 1 y 2 correlaciona con la pérdida de ARN Pol II en la región adyacente a la zona eliminada y en el exón 6, que podría contener un elemento regulatorio importante para la transcripción de *Notch* en el Dominio 2. Para esta mutante además observamos la formación de un Dominio Topológico que comprende a todo el locus de *Notch* pese a una importante disminución en la transcripción del locus. Por lo tanto, nuestros datos sugieren que, aunque la transcripción puede influir en el establecimiento de contactos al interior de los Dominios Topológicos, sitios genómicos con accesibilidad a la cromatina y ocupados por la ARN Pol II pueden tener un papel fundamental en la organización del genoma en *Drosophila* independiente de la transcripción.

La unión de Proteínas Arquitectónicas es necesaria para la actividad de frontera y puede influir en el reclutamiento de la ARN Pol II

Las Proteínas Arquitectónicas en *Drosophila* pueden mediar contactos de largo alcance sin embargo su papel en la formación de Dominios Topológicos no es claro^{37,60,81,82}. El análisis de datos de *in nucleus* Hi-C de alta resolución han llevado a la propuesta de un modelo de organización de la cromatina en el que los Dominios Topológicos emergen como una consecuencia natural de la transcripción o bien por la agregación de regiones de la cromatina con perfiles epigenéticos similares^{25,77}. Bajo este modelo, la unión de Proteínas Arquitectónicas en las fronteras de genes proporciona un nivel adicional de aislamiento, es decir, disminuyen los contactos físicos ectópicos entre dominios pero no son necesarias para la formación de Dominios Topológicos²⁵. Por otra parte, se ha propuesto que motivos de unión al ADN para Proteínas Arquitectónicas, como los motivos de M1BP y BEAF-32, o bien la unión de CP190 pueden guiar la organización de los cromosomas en Dominios Topológicos esto basado en análisis de datos de alta resolución de *in nucleus* Hi-C y modelaje^{29,83,84}. Además, la mutación de sitios de unión a Proteínas Arquitectónicas como GAF puede afectar contactos de largo alcance entre elementos regulatorios⁶⁹ lo cual apoya un papel topológico para sitios de unión a Proteínas Arquitectónicas. Por otra parte, los datos de Hi-C derivados de células con eliminación de Proteínas Arquitectónicas como BEAF-32 y M1BP no muestran datos conclusivos sobre el papel de al menos estas Proteínas Arquitectónicas en la organización del genoma²⁹ lo cual pone en duda el papel de estas proteínas en la organización del genoma en *Drosophila melanogaster*. Nuestros datos apoyan la propuesta de que los sitios de unión a Proteínas Arquitectónicas son importantes para la actividad de frontera.

La frontera 5' de *Notch* (F1) tiene un sitio de unión para dCTCF y M1BP pese a estar ocupada por más de ocho diferentes Proteínas Arquitectónicas, lo cual probablemente refleja el reclutamiento de no-directo de las otras proteínas a consecuencia de su interacción física con CP190⁸¹. Observamos que aquellas mutantes CRISPR que eliminan el sitio de unión a dCTCF (5pN- Δ 183, 5pN- Δ 343, 5pN- Δ 755) afectan la capacidad de aislamiento de la frontera F1 lo cual se acompaña no tan solo con la pérdida de unión de dCTCF y sino con la pérdida de unión de ARN Pol II y de la marca post-traducciona de histonas H3K27ac. También observamos que la mutación del sitio de unión a dCTCF en un ensayo de movilidad electroforética (EMSA) resulta en la pérdida de reclutamiento de proteínas nucleares a dicha región. En mamíferos, CTCF interacciona físicamente con la ARN Pol II⁸⁵ mientras que en *Drosophila* el co-factor de Proteínas Arquitectónicas CP190 interacciona directamente no tan solo con CTCF⁵⁷ si no también con complejos remodeladores de la cromatina y acetilasas de histonas⁸⁶ y co-localiza con sitios de ARN Pol II⁸¹, lo cual sugiere que dCTCF puede ser un factor esencial para la actividad de frontera de la región 5' de *Notch* a través de reclutar a CP190 lo cual podría influir en el remodelamiento de la cromatina y el reclutamiento de otras proteínas como ARN Pol II. En apoyo a esta propuesta, un

reporte reciente sugiere que dCTCF es importante para la actividad de frontera en el cluster de genes *Hox*⁸⁷ mientras que nosotros hemos identificado que en mamíferos, CTCF media la formación de una asa de cromatina que contiene al locus *Notch1*, el homólogo directo de *Notch* en *Drosophila* (Cerecedo-Castillo et al., en preparación).

Una observación interesante derivada de nuestros datos es que el efecto en el reclutamiento de ARN Pol II a la región 5' intergénica de *Notch* es más pronunciado en las mutantes que remueven el sitio de unión a dCTCF pero no al eliminar la Región Proximal al Promotor que contiene un sitio de unión a M1BP. Además, observamos que la pérdida de ARN Pol II es máxima cuando se eliminan tanto el sitio de unión a dCTCF como el de M1BP como se observa en la mutante *5pN-Δ343*, la cual además, presenta la máxima disminución de H3K27ac y es la única mutante en la que claramente observamos la pérdida del Dominio D1 de *Notch* y la fusión de Dominios Topológicos adyacentes, lo cual sugiere que las Proteínas Arquitectónicas contribuyen al reclutamiento de ARN Pol II y en el remodelación de la cromatina en las fronteras de dominios. Nuestras observaciones también sugieren que una frontera de dominio puede ser resistente a la pérdida de un sitio de unión a Proteínas Arquitectónicas en función de la presencia de motivos adicionales.

Finalmente, en soporte a la propiedad de las Proteínas Arquitectónicas en influir en el reclutamiento de la ARN Pol II, se ha reportado que la eliminación del factor pionero Zelda durante el desarrollo embrionario temprano resulta en la inactivación de fronteras y fusión de Dominios Topológicos⁶². Las fronteras afectadas fueron sitios que eran ocupados por Zelda y ARN Pol II pero que ante la eliminación de Zelda mostraban pérdida en el reclutamiento de ARN Pol II, sugiriendo que Proteínas Arquitectónicas con capacidad de unión al ADN de forma específica y la presencia de ARN Pol II son necesarias para el aislamiento de contactos entre Dominios Topológicos y por lo tanto para la actividad de frontera.

Mecanismos independientes de fronteras de Dominios Topológicos influyen en la organización de la cromatina en *Drosophila melanogaster*

Una observación importante de nuestros experimentos es que la inactivación de fronteras no tiene un efecto significativo en la organización de sub-dominios de *Notch*. Por ejemplo, en la mutante *5pN-Δ343* observamos la fusión del Dominio D1 de *Notch* con el Dominio Topológico de *kirre* localizado río arriba, sin embargo, los dos subdominios del Dominio D1 aún se pueden observar en los mapas de Hi-C a 1kb de resolución, aunque si presentan una menor frecuencia de interacciones al interior de cada subdominio. Este fenómeno es más evidente en la mutante para la frontera F2 en la cual se pierde el aislamiento entre los Dominios Topológicos de *Notch* pero esto no es suficiente para afectar la organización hacia el interior de cada dominio. Estas observaciones sugieren la existencia de mecanismos adicionales que favorecen la organización de la cromatina en dominios de contacto de menor tamaño. Un

mecanismo probable sería la agregación homotípica de regiones de cromatina asociadas a eurocromatina o heterocromatina como ha sido recientemente propuesto^{25,53,77}. En apoyo a esta propuesta, observamos que la marca post-traducciona de histonas H4K16ac están presentes a lo largo de todo el Dominio 2 y en uno de los dos subdominios del Dominio 1. Estos datos sugieren que las fronteras de dominios a través de reclutar Proteínas Arquitectónicas como CTCF y CP190 así como ARN Pol II, contribuyen a la formación de Dominios Topológicos al modular interacciones no específicas con regiones adyacentes mientras que mecanismos adicionales influyen en la compactación de la cromatina en dominios de contacto de menor tamaño²⁵.

Las fronteras de Dominios Topológicos influyen en la regulación de genes

La organización de la cromatina en Dominios Topológicos puede influir en la expresión de genes al modular las interacciones regulatorias entre enhancers y promotores^{20,34,75}. Nosotros encontramos que las fronteras de Dominios Topológicos son esenciales para la regulación de la expresión génica en *Drosophila melanogaster*. Por ejemplo, todas las eliminaciones en la región 5' intergénica de *Notch* afectan la transcripción de los exones contenidos en el Dominio 1 probablemente como consecuencia de la reducción en la ocupación de ARN Pol II en el extremo 5' del gen. Consistente con esta propuesta, la eliminación de la Región Proximal al Promotor abate la transcripción del Dominio D1 sin efectos significativos a nivel topológico. Por lo tanto, en esta mutante, el efecto en la transcripción se puede explicar por la reducción en ARN Pol II. Sin embargo, la pérdida de aislamiento también contribuye a los defectos transcripcionales observados en el resto de las mutantes. Por ejemplo, en las mutantes que presentan una reducción en la transcripción en el Dominio 1 de al menos el 50% (5pN- Δ 102, 5pN- Δ 343, 5pN- Δ 755), la región que contiene al inicio de la transcripción de *Notch* gana interacciones ectópicas con el Dominio Topológico de *kirre*, sugiriendo que, además de la pérdida de ARN Pol II, interacciones ectópicas entre dominios puede afectar la transcripción. En soporte a la interferencia transcripcional por contactos ectópicos a causa de defectos en aislamiento, se ha reportado que el fenotipo en el ojo característico en la mutante *fa(swb)* puede ser suprimido por la eliminación de una región localizada en el locus *rst*, el cual está localizado en la banda 3C6 justo río arriba de *Notch*⁸⁸ sugiriendo que secuencias en la banda 3C6 interfieren con la actividad de *Notch* probablemente de forma tejido-específico. Nosotros observamos que la fusión de los Dominios Topológicos de *kirre* y *Notch* resultan en la formación de un nuevo dominio cuya frontera 5' se localiza exactamente en el locus *rst* que es una región ocupada por CP190 y ARN Pol II. Además, detectamos una nueva interacción estadísticamente significativa entre *Notch* y *rst*, sugiriendo fuertemente que el establecimiento de interacciones ectópicas con secuencias río arriba puede afectar la transcripción de *Notch*.

La eliminación de la frontera F2 también resulta en defectos en la expresión de *Notch*. De forma interesante, la eliminación de la frontera F2 ocasiona la pérdida de ARN Pol II en la región adyacente a la eliminación, en el exón 6 de *Notch* y disminuye su reclutamiento en la región 5' intergénica de *Notch*, lo cual sugiere que esta frontera influye en el reclutamiento de ARN Pol II a elementos regulatorios en el locus, posiblemente por interacción física directa, lo cual es consistente con la presencia de un enhancer en dicha frontera. Por lo tanto en este caso, la reducción en la transcripción se puede deber a la pérdida de interacciones físicas entre elementos regulatorios que favorecen el reclutamiento de ARN Pol II y probablemente no refleja defectos en aislamiento. Además, la pérdida del mega-dominio a consecuencia de la eliminación de la frontera F2 afecta la transcripción de múltiples genes.

Por lo tanto, las fronteras de Dominios Topológicos pueden afectar la expresión de genes en *Drosophila*. Nuestras conclusiones son consistente con observaciones recientes en las que la eliminación de una frontera de dominio en el locus *Hox* en embriones de *Drosophila* afecta la expresión del gen y resulta en la fusión de Dominios Topológicos adyacentes. En oposición a estas observaciones, recientemente se reportó que el efecto de rearrreglos cromosomales si bien afecta la organización de los Dominios Topológicos tiene un efecto limitado en la expresión de genes, sugiriendo que mecanismos adicionales deben operar para asegurar la regulación de la expresión de genes⁸⁹, sin embargo, estas conclusiones provienen de experimentos usando cromosomas balanceadores de *Drosophila* como modelo de rearrreglos cromosomales y dificulta la interpretación de los efectos topológicos en la expresión de genes puesto que es posible que los efectos que observamos a nivel regulatorio sean adaptaciones a esos rearrreglos.

En conclusión, en este trabajo demostramos experimentalmente la existencia de secuencias genéticas discretas con actividad de frontera que influyen en la organización del genoma en Dominios Topológicos y sugiere fuertemente que regiones adicionales en el genoma de *Drosophila* podrían tener una función similar. Basados en nuestros resultados, proponemos un mecanismo para la formación de fronteras que depende de la ganancia de accesibilidad a la cromatina por la unión de Proteínas Arquitectónicas que reclutan CP190 y ARN Pol II. En este modelo, la actividad de una frontera es robusta a la pérdida de Proteínas Arquitectónicas en la medida en la que la ocupación de ARN Pol II y la accesibilidad de la cromatina se mantengan. Por lo tanto proponemos que en la organización de la cromatina al interior del núcleo en *Drosophila* es dependiente de dos mecanismos. Uno que es reflejo de la agregación de segmentos de la cromatina con perfiles transcripcionales o epigenéticos similares y otros que proporciona un nivel adicional de aislamiento que depende de la actividad de elementos genéticos que actúan como aisladores de la cromatina.

Capítulo 9.

Perspectivas

En el presente trabajo describimos los efectos estructurales y transcripcionales que se observan al eliminar las fronteras de dominios en el locus de *Notch*. Nuestro trabajo sugiere que secuencias genéticas con motivos de unión para Proteínas Arquitectónicas y que son ocupadas por la ARN Pol II son importantes para la estructuración en Dominios Topológicos en *Drosophila*. Si bien nuestro trabajo representa una importante contribución al campo, desconocen aún detalles importantes sobre la organización del genoma y su impacto en la regulación genética.

En cuanto al papel de las fronteras de dominios en la organización del genoma en *Drosophila* es necesario extender nuestro análisis a otras regiones del genoma. Actualmente contamos con métodos experimentales que nos permiten inactivar múltiples fronteras de Dominios Topológicos en una misma célula y evaluar su efecto topológico y regulatorio. Este experimento extendería nuestra observación sobre *Notch* y sugeriría que las fronteras de dominio son fundamentales para la organización del genoma.

En cuanto al papel de las Proteínas Arquitectónicas, es necesario demostrar el efecto de eliminar solo un motivo de unión a Proteínas Arquitectónicas en las fronteras de Dominio. En nuestro caso, mutar el motivo de unión a CTCF o M1BP podría brindar información fundamental sobre el papel que cada una de estas proteínas tiene en la actividad de fronteras. Adicionalmente, será fundamental realizar experimentos en los cuales se elimine de forma aguda a cada una de las Proteínas Arquitectónicas en células de *Drosophila* y evaluar su efecto topológico. Los resultados de estos experimentos darían información fundamental sobre la contribución individual de estas proteínas a la organización del genoma en *Drosophila*.

Finalmente, es aún una incógnita porque existen sitios genómicos ocupados por Proteínas Arquitectónicas o ARN Pol II que no se comportan como fronteras de dominios. Es posible que existan elementos adicionales que modulen que ciertas regiones del genoma funcionen como fronteras. La inactivación de CP190 por modificaciones postraduccionales podría ser una forma de inactivar fronteras sin afectar la ocupación de las Proteínas Arquitectónicas de unión directa al ADN. Experimentos de proteómica locus específicos podrían aportar información importante al respecto.

Bibliografía

1. Morgan, T. H. An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila*. *J. Exp. Zool.* **11**, 365–413 (1911).
2. Muller, H. J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.* **22**, 299–334 (1930).
3. Hannah, A. Localization and Function of Heterochromatin in *Drosophila Melanogaster*. in *Advances in Genetics* (ed. Demerec, M.) vol. 4 87–125 (Academic Press, 1951).
4. Cedar, H. & Felsenfeld, G. Transcription of chromatin in vitro. *J. Mol. Biol.* **77**, 237–254 (1973).
5. Kornberg, R. D. & Thomas, J. O. Chromatin structure; oligomers of the histones. *Science* **184**, 865–868 (1974).
6. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532–537 (1984).
7. Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* **65**, 1023–1031 (1991).
8. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851 (1996).
9. The genome sequence of *Drosophila melanogaster*. - PubMed - NCBI.
<https://www.ncbi.nlm.nih.gov/pubmed/10731132>.
10. The axolotl genome and the evolution of key tissue formation regulators | Nature.
<https://www.nature.com/articles/nature25458>.

11. The largest eukaryotic genome of them all? - PELLICER - 2010 - Botanical Journal of the Linnean Society - Wiley Online Library.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8339.2010.01072.x>.
12. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353 (1982).
13. Spradling, A. C. & Rubin, G. M. The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* **34**, 47–57 (1983).
14. Schwartz, Y. B. & Cavalli, G. Three-Dimensional Genome Organization and Function in *Drosophila*. *Genetics* **205**, 5–24 (2017).
15. The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in *Drosophila melanogaster* | The American Naturalist: Vol 88, No 841.
<https://www.journals.uchicago.edu/doi/abs/10.1086/281833>.
16. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306–1311 (2002).
17. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 1–20 (2019) doi:10.1038/s41576-019-0195-2.
18. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
19. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
20. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).

21. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **160**, 1145–1158 (2015).
22. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
23. Arzate-Mejía, R. G., Recillas-Targa, F. & Corces, V. G. Developing in 3D: the role of CTCF in cell differentiation. *Dev. Camb. Engl.* **145**, (2018).
24. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
25. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* (2017) doi:10.1016/j.molcel.2017.07.022.
26. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
27. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
28. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
29. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
30. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
31. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).

32. Downen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
33. Akhtar, W. *et al.* Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154**, 914–927 (2013).
34. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).
35. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
36. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).
37. Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* **15**, R82 (2014).
38. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
39. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152 (2017).
40. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28 (2017).
41. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
42. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
43. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **advance online publication**, (2017).

44. Jung, Y. H. *et al.* Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep.* **18**, 1366–1382 (2017).
45. Kaaij, L. J. T., van der Weide, R. H., Ketting, R. F. & de Wit, E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep.* **24**, 1-10.e4 (2018).
46. Ciabrelli, F. *et al.* Stable Polycomb-dependent transgenerational inheritance of chromatin states in *Drosophila*. *Nat. Genet.* (2017) doi:10.1038/ng.3848.
47. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456–E6465 (2015).
48. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
49. Larson, A. G. *et al.* Liquid droplet formation by HP1a suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
50. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, (2018).
51. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
52. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
53. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
54. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).

55. Eagen, K. P., Hartl, T. A. & Kornberg, R. D. Stable chromosome condensation revealed by chromosome conformation capture. *Cell* **163**, 934–946 (2015).
56. Cubeñas-Potts, C. & Corces, V. G. Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett.* **589**, 2923–2930 (2015).
57. Mohan, M. *et al.* The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.* **26**, 4203–4214 (2007).
58. Li, Y. *et al.* The structural basis for cohesin–CTCF-anchored loops. *Nature* 1–9 (2020) doi:10.1038/s41586-019-1910-z.
59. Pugacheva, E. M. *et al.* CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci.* (2020) doi:10.1073/pnas.1911708117.
60. Vogelmann, J. *et al.* Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the Drosophila Genome. *PLOS Genet.* **10**, e1004544 (2014).
61. Rowley, M. J. *et al.* Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization. *Cell Rep.* **26**, 2890-2903.e3 (2019).
62. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228.e19 (2017).
63. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266 (2011).
64. Andersson, E. R., Sandberg, R. & Lendahl, U. Notch signaling: simplicity in design, versatility in function. *Dev. Camb. Engl.* **138**, 3593–3612 (2011).
65. Welshons, W. J. & Keppy, D. O. Intragenic deletions and salivary band relationships in Drosophila. *Genetics* **80**, 143–155 (1975).

66. Keppy, D. O. & Welshons, W. J. The cytogenetics of a recessive visible mutant associated with a deficiency adjacent to the notch locus in *Drosophila melanogaster*. *Genetics* **85**, 497–506 (1977).
67. Rykowski, M. C., Parmelee, S. J., Agard, D. A. & Sedat, J. W. Precise determination of the molecular limits of a polytene chromosome band: regulatory sequences for the Notch gene are in the interband. *Cell* **54**, 461–472 (1988).
68. Li, L. *et al.* Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol. Cell* **58**, 216–231 (2015).
69. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Mol. Cell* **71**, 73-88.e5 (2018).
70. Chathoth, K. T. & Zabet, N. R. Chromatin architecture reorganization during neuronal cell differentiation in *Drosophila* genome. *Genome Res.* **29**, 613–625 (2019).
71. modENCODE Consortium *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
72. Simón, R., Aparicio, R., Housden, B. E., Bray, S. & Busturia, A. *Drosophila* p53 controls Notch expression and balances apoptosis and proliferation. *Apoptosis* **19**, 1430–1443 (2014).
73. Ramos, R. G., Grimwade, B. G., Wharton, K. A., Scottgale, T. N. & Artavanis-Tsakonas, S. Physical and functional definition of the *Drosophila* Notch locus by P element transformation. *Genetics* **123**, 337–348 (1989).
74. Vazquez, J. & Schedl, P. Deletion of an insulator element by the mutation facet-strawberry in *Drosophila melanogaster*. *Genetics* **155**, 1297–1311 (2000).
75. Despang, A. *et al.* Functional dissection of TADs reveals non-essential and instructive roles in regulating gene expression. *bioRxiv* 566562 (2019) doi:10.1101/566562.

76. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
77. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).
78. Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
79. Stadler, M. R., Haines, J. E. & Eisen, M. B. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo. *eLife* **6**, e29550 (2017).
80. Andreenkov, O. V., Volkova, E. I., Semeshin, V. F., Zhimulev, I. F. & Demakov, S. A. Structural features of 3C6/C7 interband chromatin organization in *Drosophila melanogaster* polytene chromosomes. *Cell Tissue Biol.* **7**, 347–351 (2013).
81. Liang, J. *et al.* Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing. *Mol. Cell* **53**, 672–681 (2014).
82. Wood, A. M. *et al.* Regulation of Chromatin Organization and Inducible Gene Expression by a *Drosophila* Insulator. *Mol. Cell* **44**, 29–38 (2011).
83. Mourad, R. & Cuvier, O. Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. *PLoS Comput. Biol.* **12**, e1004908 (2016).
84. Mourad, R. & Cuvier, O. TAD-free analysis of architectural proteins and insulators. *Nucleic Acids Res.* **46**, e27 (2018).
85. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
86. Ali, T. *et al.* Chromatin binding of Gcn5 in *Drosophila* is largely mediated by CP190. *Nucleic Acids Res.* **45**, 2384–2395 (2017).

87. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
88. Welshons, W. J. & Welshons, H. J. Suppression of the facet-strawberry position effect in *Drosophila* by lesions adjacent to notch. *Genetics* **110**, 465–477 (1985).
89. Ghavi-Helm, Y. *et al.* Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **1** (2019) doi:10.1038/s41588-019-0462-3.
90. Gerasimova, T. I., Lei, E. P., Bushey, A. M. & Corces, V. G. Coordinated Control of dCTCF and gypsy Chromatin Insulators in *Drosophila*. *Mol. Cell* **28**, 761–772 (2007).
91. Haeussler, M. *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
92. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Mutagenesis and homologous recombination in *Drosophila* cell lines using CRISPR/Cas9. *Biol. Open* **3**, 42–49 (2014).
93. Böttcher, R. *et al.* Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Res.* **42**, e89 (2014).
94. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).
95. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
96. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
97. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).

98. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
99. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
100. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
101. Stansfield, J. C., Cresswell, K. G., Vladimirov, V. I. & Dozmorov, M. G. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinformatics* **19**, 279 (2018).
102. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
103. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
104. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).
105. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).
106. Meier, K. *et al.* LINT, a Novel dL(3)mbt-Containing Complex, Represses Malignant Brain Tumour Signature Genes. *PLOS Genet.* **8**, e1002676 (2012).
107. García-González, E. & Recillas-Targa, F. A regulatory element affects the activity and chromatin structure of the chicken α -globin 3' enhancer. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1839**, 1233–1241 (2014).

108. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
109. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

Anexo I.

Métodos

Cultivo de células

Para los experimentos presentados en este trabajo se usaron células S2R+. Estas células fueron cultivadas a 25 ° C en medio Schneider (Thermofisher) suplementado con 10% de Suero Fetal Bovino (Biowest) y penicilina/estreptomicina (Invitrogen).

Anticuerpos

Los siguientes anticuerpos se usaron para ChIP: el anticuerpo policlonal para dCTCF fue generado por New England Peptide inmunizando conejos con un péptido correspondiente a los primeros 20 aminoácidos de dCTCF⁹⁰. Los siguientes anticuerpos se obtuvieron de fuentes comerciales: anti-H3K4me3 (Abcam # 8580), anti-H3K27ac (Abcam # 4729), anti-H3K27me3 (Abcam # 6002), anti-RNA Pol II pSer2 (Abcam # 5095), anti-RNA Pol II pSer5 (Abcam # 5408), ratón anti-IgG (Millipore # 12-371), conejo anti-IgG (Millipore # 12-370).

CRISPR-Cas9

Diseño y clonación de sgRNAs en el plásmido pAc-sgRNA-Cas9. Para el diseño y evaluación de secuencias guía para el sistema CRISPR-Cas9, utilizamos el software CRISPROR (<http://crispor.tefor.net/>)⁹¹ y la versión dm6 del genoma de *Drosophila melanogaster*. Las regiones genómicas utilizadas para el diseño de los guías fueron chrX: 3,134,000-3,134,869 y chrX: 3,160,210-3,163,651 que corresponden a la región 5 'de *Notch* y al enhancer intrónico, respectivamente. Las secuencias guía seleccionadas tenían una de las puntuaciones más bajas de off-targets. Las secuencias para todas las guías utilizadas en este estudio se proporcionan en la Tabla 2. Es importante destacar que las guías utilizadas para generar eliminaciones en la región 5 'de *Notch* no reconocen a la secuencia que contiene los principales sitios del inicio de la transcripción anotados para *Notch*, mientras que las guías eliminar el enhancer intrónico se ubican al menos a 100 pb del exón más cercano. Los oligonucleótidos para guías fueron adquiridos a través de Sigma y se clonaron en el plásmido pAc-sgRNA-Cas9 (Addgene # 49330)⁹². La secuencia de las guías clonadas fue evaluada por Secuenciación de Sanger. Los plásmidos se expandieron y purificaron usando un kit Qiagen MiniPrep antes de la transfección.

Transfecciones. Para generar eliminaciones de segmentos genómicos de interés empelamos una estrategia de dos guías mediante la cual se dirige un guía a cada uno de los extremos de la región a eliminar (ver Figura 16d y Figura Suplementaria 17). Se

realizaron transfecciones en células S2R + como se describe en Basset et al., 2014⁹² con modificaciones importantes. En resumen, 600,00 células S2R + se sembraron en cada pozo de una caja de 24 pozos resuspendidas en 500 µl de medio Schneider. Para la transfección usamos FugeneHD (Promega) siguiendo las recomendaciones del fabricante (<https://worldwide.promega.com/techserv/tools/FugeneHdTool/>). Se usó una relación 4:2 FugeneHD:plásmido para todas las transfecciones. Las células transfectadas se colocaron a 25°C y 72 horas después de la transfección, las células se resuspendieron en medio fresco suplementado con 5µg/ml de Puromicina (Sigma) y se transfirieron a un pocillo de una placa de 12 pocillos. Después de 3 días (6 días después de la transfección), las células se volvieron a suspender en medio nuevo suplementado con 5µg/ml de Puromicina (Sigma) y se transfirieron a un pocillo de una placa de 6 pocillos. Después de 3 días (9 días después de la transfección) se usó una alícuota de células para extracción de ADN mediante fenol-cloroformo y genotipificación por PCR usando oligonucleótidos específicos que abarcan las eliminaciones deseadas (ver Tabla 2). Las poblaciones de células mutantes se expandieron en medio Schneider fresco sin puromicina y se usaron para el protocolo de aislamiento de clones mutantes.

Aislamiento clonal de células mutantes. Se usaron poblaciones de células mutantes para dilución serial como se ha descrito con modificaciones menores⁹³. Las células mutantes se resuspendieron en medio condicionado (30% de medio de crecimiento de células mutantes durante 72 horas y 70% de medio de Schneider fresco con 10% de FBS y penicilina / estreptomina) y se añadieron 100 ul de medio con células por pocillo en una placa de 96 pocillos. Las células se incubaron a 25°C durante al menos 3 semanas. Los clones individuales se expandieron y se usó una alícuota de células para extracción de ADN mediante fenol-cloroformo y genotipificación por PCR usando oligonucleótidos específicos que abarcan las eliminaciones deseadas (Tabla 2). Los clones homocigotos se expandieron y se usaron para experimentos posteriores. Los clones mutantes se caracterizaron adicionalmente por secuenciación de Sanger. Los fragmentos amplificados se ligaron en pGEM-T Easy y se usaron dos clones individuales para la secuenciación de Sanger por mutante. La secuencia de cada fragmento mutante fue la misma entre clones para todos los mutantes CRISPR generados en este estudio.

Generación y procesamiento de datos *in nucleus* Hi-C

***in nucleus* Hi-C y secuenciación.** Generamos bibliotecas de Hi-C para células S2R+ control y mutantes de CRISPR por duplicado usando un protocolo de *in nucleus* Hi-C con modificaciones menores para células de *Drosophila*⁹⁴. Brevemente, 40x10⁶ células se fijaron con formaldehído a una concentración final de 2% durante 10 minutos a temperatura ambiente y la reacción de fijación se detuvo agregando glicina 1M durante 5 minutos. Las células se lavaron con PBS frío y se congelaron en nitrógeno líquido para su almacenamiento a corto plazo a -70 ° C o se usaron inmediatamente para experimentos posteriores. Los genomas se digirieron con Mbol (NEB) y los extremos

cohesivos se repararon con nucleótidos y Biotin-14-dATP (Invitrogen). Las bibliotecas Hi-C se secuenciaron en una plataforma HiSeq 2500 de 50 pb en modo Pair-End.

Generación de genomas mutantes *in silico*. Puesto que las mutantes CRISPR tienen eliminaciones de fragmentos del genoma es necesario hacer el mapeo de lecturas en genomas sintéticos que contengan dichas eliminaciones. Se construyeron genomas mutantes *in silico* para todos los mutantes CRISPR. La secuencia de cada mutante determinada por la secuenciación de Sanger se utilizó para reemplazar la secuencia de tipo silvestre en el cromosoma X usando R. Se construyó un índice específico para cada genoma mutante usando Bowtie2⁹⁵. Además, se utilizaron genomas mutantes para generar archivos de digestión Mbol específicos de mutantes para su posterior análisis utilizando HiC-Pro `digest_genome.py`.

Procesamiento de datos. El mapeo, el filtrado, la corrección y la generación de matrices Hi-C se realizaron utilizando HiC-Pro⁹⁶. Brevemente, cada par de lecturas de secuenciación se mapearon independientemente a los genomas mutantes o de tipo silvestre usando Bowtie2 con los parámetros predeterminados de HiC-Pro. Después del filtrado, se usaron pares válidos para generar matrices de interacción sin procesar o normalizadas por ICE a una resolución de 1 kb, 5 kb, 20 kb y 100 kb. Para corregir las diferencias en la profundidad de secuenciación entre experimentos, las matrices se normalizaron por el menor número de pares de Hi-C válidos (Tabla 1). Todos los mapas de calor de matrices de contacto normalizadas se generaron con HiCPlotter⁹⁷. Las gráficas de correlación entre todas las réplicas de Hi-C y las gráficas de conteo vs distancia se generaron usando HiCExplorer²⁹ y matrices con un tamaño de bin de 10 kb.

Identificación de TADs. Los TAD se identificaron para matrices con una resolución de 1 kb y 5 kb para mutantes CRISPR y silvestre usando el TAD separation score de HiCExplorer. La identificación de TADs a 1 kb de resolución se realizó con los parámetros `-minDepth 10000 -maxDepth 40000 -step 1500 -thresholdComparisons 0.0000001 -delta 0.01` y `-correctForMultipleTesting fdr`. La identificación de TADs a una resolución de 5 kb se realizó con los parámetros `-minDepth 20000 -maxDepth 200000 -step 10000 -thresholdComparisons 0.01 -delta 0.01` y `-correctForMultipleTesting fdr`. Se generaron BoxPlots de recuentos de Hi-C para interacciones intra e interdominio en R usando matrices normalizadas de HiC-Pro y las coordenadas de TADs para la muestra de tipo silvestre obtenidas de HiCExplorer. Se usó una prueba de suma de rangos de Wilcoxon para determinar diferencias estadísticamente significativas.

Análisis virtual 4C. Los perfiles de 4C Virtuales para diferentes viewpoints se generaron utilizando HiC-Pro `make_viewpoints.py` y el mismo número de pares válidos para los conjuntos de datos de interés. IGV se utilizó para mostrar las interacciones asociadas a cada viewpoint. El porcentaje de pares válidos con el viewpoint de interés en una región genómica específica se calculó utilizando R (<http://www.R-project.org>).

Análisis de compartimientos. Los compartimientos se identificaron usando HOMER⁹⁸ runHiCpca.pl con los parámetros `-res 10000 -window 50000`. Los picos de enriquecido para la marca de histona H3K4me3 en las células S2R + (GSM2259985) se identificaron usando MACS2⁹⁹ y se usaron para asignar el signo a los resultados de PCA1. Los resultados del PCA1 se graficaron usando IGV¹⁰⁰.

Análisis diferencial de conjuntos de datos Hi-C. Se usó HiCcompare¹⁰¹ con los parámetros predeterminados para identificar diferencias en las interacciones de cromatina entre condiciones. Se realizaron comparaciones por pares entre mutantes de tipo salvaje y CRISPR utilizando matrices de contacto sin normalizar a una resolución de 5 y 20 kb. Los contactos diferenciales en las regiones de interés se graficaron utilizando el Epigenome Browser de la Universidad de Washington.

Procesamiento de datos Hi-C de conjuntos de datos públicos. Los conjuntos de datos Hi-C públicos de embriones de *Drosophila* se obtuvieron de la base de datos del Archivo Europeo de Nucleótidos (ENA) con los siguientes números de acceso: embriones del ciclo nuclear 12 (ERR1533189-199), embriones del ciclo nuclear 13 (ERR1533200-209), embriones del ciclo nuclear 14 (ERR1533226-236), embriones de 3-4 horas (ERR1533170-181), embriones del ciclo nuclear 14 tratamiento con triptolida (ERR1912894-899). Los datos se procesaron de la misma manera que para las células S2R+ utilizando Bowtie 2 y HiC-Pro. Los mapas de calor de las matrices de contacto se generaron a partir de HiCPlotter.

Generación y procesamiento de datos de RNA-seq

El ARN total se extrajo de la mutantes CRISPR para el enhancer de la frontera F2 y de células S2R+ silvestre por duplicado usando TRIzol (ThermoFisher) siguiendo las instrucciones del fabricante. Las bibliotecas de ARN totales para todas las muestras fueron generadas y secuenciadas por Novogene. Se secuenciaron dos bibliotecas independientes por condición en una plataforma HiSeq 4000 de 150 pb. Los datos de RNA-seq se analizaron usando Salmon¹⁰² y DESeq2¹⁰³. Brevemente, estimamos la abundancia de cada transcrito usando Salmon con un índice específico para el transcriptoma *Drosophila melanogaster* (dm6) y las opciones `-validateMappings` y `--gcBias`. Los datos de cuantificación de Salmon se importaron a R y creamos una tabla con datos de cuantificación a nivel de gen como input para DESeq2. Calculamos la expresión diferencial (DE) entre condiciones usando DESeq2. Determinamos que un gen se encontraba diferencialmente expresado si presentaba una razón de cambio de al menos $\log_2(0.5)$ y un valor *p* corregido de <0.05 . Los niveles de abundancia para los genes ubicados en el Mega-dominio se transformaron en un valor Z y se usaron para crear un mapa de calor usando R. Para crear tracks de señal a partir de datos de RNA-seq, se mapearon las lecturas de secuenciación contra el genoma de *Drosophila melanogaster* (dm6) usando Bowtie2. Los archivos SAM se convirtieron en archivos BAM

utilizando samtools. Los archivos BAM se ordenaron y se indexaron. Se utilizó deepTools v3.3.0¹⁰⁴ para calcular la correlación de Pearson y PCA para todos los conjuntos de datos. Los archivos BAM de las réplicas se fusionaron debido a la alta correlación (Correlación de Pearson > 0.9) y los archivos BAM de los experimentos combinados se usaron para crear tracks de señal con bamCoverage –normalizeUsing BPM. Las track de señal para los datos de RNA-seq de mutantes y la condición silvestre se visualizaron usando IGV.

RT-PCR

El ARN total se extrajo de células de tipo silvestre y mutantes CRISPR al menos por duplicado usando Trizol y siguiendo las instrucciones del fabricante. El ARN se usó directamente para qRT-PCR utilizando el kit KAPA SYBR FAST One Step (KAPA Biosystems) en al menos dos réplicas técnicas por muestra utilizando un sistema de PCR en tiempo real StepOne. En la tabla 2 se proporciona una lista de todos los conjuntos de oligonucleótido utilizados en este estudio. Cuando fue posible, los oligonucleótidos se diseñaron para abarcar uniones exón-exón. El gen RP49 expresado constitutivamente se usó como control. Los datos se analizaron mediante el método $\Delta\Delta C_t$ ¹⁰⁵. Las diferencias estadísticamente significativas en la expresión génica entre condiciones se calcularon utilizando una prueba t (valor $p < 0,05$) y el software Graphpad Prisma 7.0.

ChIP y qPCR

La inmunoprecipitación de cromatina usando células de tipo silvestre y todos los mutantes CRISPR se realizó por duplicado como se describió previamente¹⁰⁶. Se usó ADN purificado para la cuantificación específica de las regiones de interés mediante qPCR usando SYBR Green en duplicado. Los valores medios para todas las regiones analizadas en diferentes condiciones se expresaron como enriquecimiento en comparación con IgG. Las diferencias estadísticamente significativas entre las muestras se calcularon usando una prueba t (valor $p < 0,05$) y el software Graphpad Prisma 7.0. Los oligonucleótidos para chip se muestran en la tabla 2.

EMSA

El ensayo de cambio de movilidad electroforética (EMSA) se realizó como se describió previamente con modificaciones menores¹⁰⁷. En particular, la proteína nuclear se extrajo usando ~ 40x10⁶ células S2R+. Para los ensayos de retardo se usaron 2 ug de a-dCTCF o anticuerpo de conejo IgG. Las secuencias de oligonucleótidos utilizadas para los ensayos de EMSA se muestran en la Tabla 2.

ChIP-seq y ATAC-seq recuperación y procesamiento de datos

Los datos de ChIP-seq disponibles públicamente para células S2 / S2R + y embriones de *Drosophila* se obtuvieron de la base de datos Gene Expression Omnibus (GEO). Los datos sin procesar para las células S2/S2R + se obtuvieron con los siguientes números de

acceso: CP190 (GSM1015404), SuHw (GSM1015406), Mod (GSM1015408), CTCF (GSM1015410), Ibf1 (GSM1133264), Ibf2 (GSM1133265), BEAF32 (GSM1278639), Pita (GSM1313420), ZIPIC (GSM1313421), RNA Pol II (GSM2259975), H3K4me1 (GSM2259983), H3K4me3 (GSM2259985), H3K27ac (GSM2259987), MSL2 (GSM2469507), H4K16ac (GSM246950P), GSM (701)), Entrada (GSM1015412). Los datos sin procesar para los embriones de *Drosophila* se obtuvieron con los siguientes números de acceso: Zelda nc8 (GSM763060), Zelda nc13 (GSM763061), Zelda nc14 (GSM763062), TBP MBT (GSM1022898, GSM1022899), TBP post-MBT (GSM1022903, GSM1022911), TBP pre-MBT (GSM1022912), RNA Pol II-pSer5-nc12 (GSM1536376), RNA Pol II-pSer5-nc13 (GSM1536379), RNA Pol II-pSer5-nc14early (GSM1536382), RNA Pol II-pSer5-nc14mid (GSM1536384), RNA Pol II-pSer5-nc14late (GSM1536386), RNA Pol II-pSer5-nc13ZeldaKD (GSM1536390). Los datos brutos de los embriones de *Drosophila* se obtuvieron de ENA con los siguientes números de acceso: RNA Pol II-pan nc14 (ERR1912880, ERR1912881), RNA Pol II-pSer5 nc14 (ERR1912882, ERR1912883), RNA Pol II-pan nc14-triptolida (ERR1912890, ERR1912891), ARN Pol II-pSer5 nc14-triptolida (ERR1912892, ERR1912893). Los datos ATAC-seq de los embriones de *Drosophila* se obtuvieron de GEO con los siguientes números de acceso: ATAC-nc11-3hrs (GSM2219678), ATAC-nc11-6hrs (GSM2219681), ATAC-nc11-9hrs (GSM2219684), ATAC-nc12-3hrs (GSM2219687), ATAC-nc12-6hrs (GSM2219690), ATAC-nc12-9hrs (GSM2219693), ATAC-nc12-12hrs (GSM2219696), ATAC-nc13-3hrs (GSM2219700), ATAC-nc13-6hrs (GSM2219703), ATAC-nc13-9hrs (GSM2219706), ATAC-nc13-12hrs (GSM2219709), ATAC-nc13-15hrs (GSM2219712), ATAC-nc13-18hrs (GSM2219715).

Las lecturas se mapearon contra el genoma de *Drosophila melanogaster* (dm6) usando Bowtie2 con parámetros predeterminados para lecturas individuales y pareadas. Las lecturas mapeadas se filtraron por la calidad del mapeo (-q 30) usando samtools. Los archivos BAM fueron ordenados e indexados. Los duplicados se eliminaron con Picard. Los archivos BAM se importaron a deeTools v3.3.1 para crear tracks de señal con bamCoverage -normalizeUsing CPM. La track de señal para todos los datos se visualizaron usando IGV.

Análisis de motivos

La identificación de los sitios de unión para las proteínas arquitectónicas se realizó utilizando FIMO¹⁰⁸ y la base de datos de insectos JASPAR¹⁰⁹ utilizando como valor de corte un valor $p < 00001$

Disponibilidad de datos

Los datos sin procesar para los experimentos de núcleo Hi-C y RNA-seq generados en este estudio se han depositado en el Gene Expression Omnibus (GEO) como una SuperSerie con el número de acceso: GSE136137 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136137>].

Tabla 1. Estadísticas de experimentos de Hi-C

	wild-type	5pN-Δ102	5pN-Δ183	5pN-Δ343	5pN-Δ755	Δenhancer
valid_interaction	73,371,778	80,145,697	65,248,465	74,781,396	68,610,520	68,348,417
valid_interaction_rmdup	72,953,774	79,091,829	64,026,434	73,588,490	67,582,002	67,252,940
trans_interaction	4,426,563	3,114,910	5,925,899	2,640,130	2,462,914	5,104,348
%trans_interaction	6.06762715	3.93834615	9.25539442	3.58769422	3.64433418	7.58977674
cis_interaction	68,527,211	75,976,919	58,100,535	70,948,360	65,119,088	62,148,592
%cis_interaction	93.9323728	96.0616539	90.7446056	96.4123058	96.3556658	92.4102233
cis_shortRange	33,897,767	39,009,711	28,353,246	37,199,484	34,714,319	32,296,687
cis_longRange	34,629,444	36,967,208	29,747,289	33,748,876	30,404,769	29,851,905

Table 2. Oligonucleótidos empleados en este estudio**sgRNAs**

Nombre	Secuencia 5'-3'	Uso	Notas
sgRNA-1	GGCTAACGTTATTTGTTAC	sgRNA	102-5p, 183-3p and 755-3p
sgRNA-2	ACGCGGTCACACTGCCGATT	sgRNA	102-3p and 343-3p
sgRNA-3	GTGCCTACGAATTTTACATT	sgRNA	183-5p and 343-5p
sgRNA-4	TTATGATTCCTCGTTGGGT	sgRNA	755-5p

Genotipificación

Nombre	Secuencia 5'-3'	Uso	Notas
N5p_F	GCATAAATGTGTATGTCAACGCT	genotipificación	CRISPR 5p N mutants
N5p_R	CGGCTTCGTCTCACTCTCA	genotipificación	
enh_F	CGCAATTTTCGCCGAGATT	genotipificación	CRISPR enhancer mutants
enh_R	CAGTCTTGCCCTCGAAACC	genotipificación	

ChIP

Nombre	Secuencia 5'-3'	Uso	Notas
AbdB_F	CCGCCCTTTTCGAATACAG	ChIP	control
AbdB_R	TGCACCCTCCTAGTCTAGTG	ChIP	
5pN_F	AGCGTAATTCTACATACCGCT	ChIP	ChIP for 5p end of N
5pN_R	TGGTGAACAAATAACGTTAGCC	ChIP	
enh_F	ACTTTGTTGCCCATTCGCT	ChIP	ChIP for enhancer of N
enh_R	CCAACTGCGGCCACAATTAT	ChIP	

prom_exon6_F	TCAAGTGCCTCTGTGATCCC	ChIP	ChIP for exonic promoter of N
prom_exon6_R	AATAGTAGCTGCCACACGA	ChIP	
Δ102_F	TACCGCTATGACGGCACTAA	ChIP	ChIP for CRISPR mutant
Δ102_R	GCTTTCACAACCGTTTCGTC	ChIP	
Δ183_F	TGCTACAAGTGCCTTTTCAA	ChIP	ChIP for CRISPR mutant
Δ183_R	ACAACCGTTTCGTCCTG	ChIP	
Δ343_F	CGTATTGCGGGAAAACCTAA	ChIP	ChIP for CRISPR mutant
Δ343_R	GCTTTCACAACCGTTTCGTC	ChIP	
Δ755_F	AGAAAAACTTAAAAAGCTTAGAAACAA	ChIP	ChIP for CRISPR mutant
Δ755_R	CGTCGGTGCAAGAAAAAGA	ChIP	
Δenh_F	TCTATTGCAACGTGATGTCCA	ChIP	ChIP for CRISPR mutant
Δenh_R	GGGTAAAACGGTCACACTTCC	ChIP	

RT-PCR

Nombre	Sequencia 5'-3'	Uso	Notas
rst_F	TGCACGGTGGTCAACGATTA	RT-PCR	kirre TAD
rst_R	CCAAAATGGTCAGCACCAGC	RT-PCR	
kirre_med_F	GCAATTGAACTGGCCAACGA	RT-PCR	kirre TAD
kirre_med_R	CCAAAGTTTGCTGACCGACG	RT-PCR	
kirre_full_F	TCGCGGTAATTTTGCCTTGA	RT-PCR	kirre TAD
kirre_full_R	TGGCGGCAACCTCAATTCTA	RT-PCR	
Vap33-A_F	TCGATCAACCCAGGTGGAGA	RT-PCR	N-enhancer-megadomain
Vap33-A_R	GCGGTGGGCATCTCGAAAAC	RT-PCR	
Iva_F	GGATATACGCTTAGCGGGCA	RT-PCR	N-enhancer-megadomain
Iva_R	TGGACTTGTGTCGCTCATCC	RT-PCR	
CG18508_F	CCGGAGCCAGTAAGCTAACATT	RT-PCR	N-enhancer-megadomain
CG18508_R	ACAGGATCGGCTGCACAAAT	RT-PCR	
CG6428_F	CCAGGATGTCAAAGAGGCC	RT-PCR	N-enhancer-megadomain
CG6428_R	GGCATTGGGAATGGGAGCTA	RT-PCR	
CR43298_F	AAAGGGTTGGCGTAGGGTT	RT-PCR	N-enhancer-megadomain
CR43298_R	TGAAGGACATTCACCGGACG	RT-PCR	
CR45482_F	AAGCTAAATGGCCCGTGAA	RT-PCR	N-enhancer-megadomain

CR45482_R	AAGCCCATTCTGAACCGAGAG	RT-PCR	N-enhancer- megadomain
CG43689_F	CCAGGGCATAAAGAAGCCCA	RT-PCR	
CG43689_R	ATTGAGACCAAGTCCCACGC	RT-PCR	
N_exon1_intron1_F	CGGAACTGGTGAGTAAAGCG	RT-PCR	Notch Domain 1
N_exon1_intron1_R	AATTCCAGTTTGTGCACCCA	RT-PCR	
N_exon1-exon2-F	TAACAAAATGCACGCCGTTG	RT-PCR	Notch Domain 1
N_exon1-exon2-R	ATTCTGGCAACCGACACTTG	RT-PCR	
N_exon4-exon5_F	ACACCTGTTCTATGACATCGA	RT-PCR	Notch Domain 1
N_exon4-exon5_R	TTGGTGTGCGAAATCCTTTCC	RT-PCR	
N_exon5-exon6_F	GGAATTTGCCGGTCTGAACG	RT-PCR	Notch Domain 1
N_exon5-exon6_R	GGTGTAGTCCGAGATGCCAT	RT-PCR	
N_exon8-exon9_F	AGGCCTGGAGTTCGGTTC	RT-PCR	Notch Domain 2
N_exon8-exon9_R	AGCCGGACATTGAACTTTGT	RT-PCR	
Rp49-F	AGCATAACAGGCCCAAGATCG	RT-PCR	Endogenous control
Rp49-R	TGTTGTCGATACCCTTGGGC	RT-PCR	
dnc_F	GGAGCAAAACTCGAGCGGTA	RT-PCR	dnc TAD
dnc_R	CCATTTTCTACATCGAAAGGCCGA	RT-PCR	

EMSA

Nombre	Secuencia 5'-3'	Uso	Notas
Fab8_sense	AAATTTCCACATTCCCGCCTTGCAGCGCCACCTGGCCTTGGTAATGTAGAAC TAGGAAGG	EMSA	Control for CTCF binding
Fab8_antisense	CCTTCCTAGTTCTACATTACCAAGGCCAGGTGGCGCTGCAAGCGGGAAT GTGGAAATT	EMSA	
N_ctcf_wt_sense	CATACCGCTATGACGGCACTAAAGCGCCATTCGGCGAAATGGGGAAACTA CTCATGCAAG	EMSA	wild-type sequence
N_ctcf_wt_antisense	CTTGATGAGTAGTTTCCCCATTCGCCGAATGGCGCTTATGCTCCATAG CGGTATG	EMSA	
N_ctcf_mut1_sense	CATACCGCTATGACGGCACTAAAGTATTATTGGCAAAATGGGGAAACTACT CATGCAAG	EMSA	mutant sequence for CTCF
N_ctcf_mut1_antisense	CTTGATGAGTAGTTTCCCCATTTGCCAAATAATACTTATGCTCCATAGCG GTATG	EMSA	
N_ctcf_mut2_sense	CATACCGCTATGACGGCACTAAAGCGCCATCAATAAAATGGGGAAACTAC TCATGCAAG	EMSA	mutant sequence for CTCF
N_ctcf_mut2_antisense	CTTGATGAGTAGTTTCCCCATTTATTGAATGGCGCTTATGCTCCATAGCG GTATG	EMSA	

Anexo II.

Publicaciones generadas durante el doctorado

Arzate-Mejía, RG, Josué Cerecedo-Castillo, A., Guerrero, G. *et al.* In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila. **Nat Commun** **11**, 894 **2020**. <https://doi.org/10.1038/s41467-020-14651-z>

Cortés-Fernández de Lara J, **Arzate-Mejía RG** and Recillas-Targa F. Enhancer-RNAs: insights into their biological role. **Epigenetic Insights**, **2019**. Review.

Carmona-Aldana F, Zampedri C, Suaste-Olmos F, Murillo-de-Ozores A, Guerrero G, **Arzate-Mejía R**, Maldonado E, Navarro RE, Chimal-Monroy J, Recillas-Targa F. CTCF knockout reveals an essential role for this protein during the zebrafish development. **Mech Dev.** **2018** May 1. pii: S0925-4773(18)30011-X.

Arzate-Mejía RG, Recillas-Targa F, Corces VG. Developing in 3D: the role of CTCF in cell differentiation. **Development.** **2018** Mar 22;145(6). Review.

Arzate-Mejía RG, Licon-Limón P, Recillas-Targa F. Genome Editing During Development Using the CRISPR-Cas Technology. **Methods Mol Biol.** **2018**;1752:177-190.

Rodríguez-Aguilera JR, Guerrero-Hernández C, Pérez-Molina R, Cadena-Del-Castillo CE, Pérez-Cabeza de Vaca R, Guerrero-Celis N, Domínguez-López M, Murillo-de-Ozores AR, **Arzate-Mejía R**, Recillas-Targa F, Chagoya de Sánchez V. Epigenetic Effects of an Adenosine Derivative in a Wistar Rat Model of Liver Cirrhosis. **J Cell Biochem.** **2018** Jan;119(1):401-413.

García-González E, Escamilla-Del-Arenal M, **Arzate-Mejía R**, Recillas-Targa F. Chromatin remodeling effects on enhancer activity. **Cell Mol Life Sci.** **2016** Aug;73(15):2897-910. Review.

Ayala-Ortega E, **Arzate-Mejía R**, Pérez-Molina R, González-Buendía E, Meier K, Guerrero G, Recillas-Targa F. Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines. **BMC Cancer.** **2016** Mar 16;16:226.

Arriaga-Canon C, Fonseca-Guzmán Y, Valdes-Quezada C, **Arzate-Mejía R**, Guerrero G, Recillas-Targa F. A long non-coding RNA promotes full activation of adult gene expression in the chicken α -globin domain. **Epigenetics.** **2014** Jan;9(1):173-81.

Arzate-Mejía RG, Valle-García D, Recillas-Targa F. Signaling epigenetics: novel insights on cell signaling and epigenetic regulation. **IUBMB Life.** **2011** Oct;63(10):881-95. Review.

“A veces, el destino se parece a una pequeña tempestad de arena que cambia de dirección sin cesar. Tú cambias de rumbo intentando evitarla. Y entonces la tormenta también cambia de dirección, siguiéndote a ti. Tú vuelves a cambiar de rumbo. Y la tormenta vuelve a cambiar de dirección, como antes. Y esto se repite una y otra vez. Como una danza macabra con la Muerte antes del amanecer. Y la razón es que la tormenta no es algo que venga de lejos y que no guarde relación contigo. Esta tormenta, en definitiva, eres tú. Es algo que se encuentra en tu interior. Lo único que puedes hacer es resignarte, meterte en ella de cabeza, taparte con fuerza los ojos y las orejas para que no se te llenen de arena e ir atravesándola paso a paso. Y en su interior no hay sol, ni luna, ni dirección, a veces ni siquiera existe el tiempo. Allí sólo hay una arena blanca y fina, como polvo de huesos, danzando en lo alto del cielo. Imagínate una tormenta como ésta.

Y cuando la tormenta de arena haya pasado, tú no comprenderás cómo has logrado cruzarla con vida. ¡No! Ni siquiera estarás seguro de que la tormenta haya cesado de verdad. Pero una cosa sí quedará clara. Y es que la persona que surja de la tormenta no será la misma persona que penetró en ella. Y ahí estriba el significado de la tormenta de arena.”

*Kafka en la orilla
Haruki Murakami
2006*

