



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

“PARTICIPACIÓN DEL FACTOR NUCLEAR CTCF Y LAS DESMETILASAS DE DNA EN LA ACTIVIDAD TRANSCRIPCIONAL DEL LCR DEL VIRUS DEL PAPILOMA HUMANO TIPO 18”

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:

MARCO ANTONIO MERAZ RODRÍGUEZ

TUTOR PRINCIPAL

DR. ERNESTO SOTO REYES SOLÍS
[Universidad Autónoma Metropolitana](#)

MIEMBROS DEL COMITÉ TUTOR

DR. ERNESTO SOTO REYES SOLÍS
[Universidad Autónoma Metropolitana](#)

DR. ALEJANDRO GARCÍA CARRANCÁ
[Instituto de Investigaciones Biomédicas](#)

DRA. LORENA AGUILAR ARNAL
[Instituto de Investigaciones Biomédicas](#)

Ciudad de México. agosto, 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.

AGRADECIMIENTOS INSTITUCIONALES

Al posgrado en Ciencias Bioquímicas de la Universidad Nacional Autónoma de México.

El presente proyecto fue realizado en el Laboratorio de Virus y Cáncer, a cargo del Dr. Alejandro García Carrancá, de la Unidad Periférica de Investigación Biomédica en Cáncer del Instituto de Investigaciones Biomédicas y el Instituto Nacional de Cancerología, y en el laboratorio de Biología Celular a cargo del Dr. Ernesto Soto Reyes Solís, del Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Este trabajo fue apoyado por los donativos del Consejo Nacional de Ciencia y Tecnología (CONACyT) Fondo CB-SEP-CONACyT (284748), UAM-PTC-704 y Desarrollo Tecnológico e Innovación en Salud ante la Contingencia por COVID-19 (312021).

Los estudios de maestría del alumno Marco Antonio Meraz Rodríguez (número de becario 630695 y número de apoyo 481908) se realizaron con el apoyo de la Beca Nacional CONACyT para posgrados.

ÍNDICE

1. RESUMEN
2. ABSTRACT
3. ANTECEDENTES
 - 3.1. CÁNCER CERVICOUTERINO
 - 3.2. ETIOLOGÍA DEL CÁNCER CERVICOUTERINO
 - 3.3. HPV Y CÁNCER CERVICOUTERINO
 - 3.4. ESTRUCTURA Y ORGANIZACIÓN DEL GENOMA DE HPV
 - 3.5. EL GENOMA CELULAR, LA CROMATINA, SU ESTRUCTURA Y REGULACIÓN
 - 3.6. INTEGRACIÓN DEL GENOMA DEL HPV AL GENOMA CELULAR
 - 3.7. ELEMENTOS GENÉTICOS QUE REGULAN LA TRANSCRIPCIÓN DEL GENOMA CELULAR
 - 3.8. ELEMENTOS DE REGULACIÓN QUE PARTICIPAN EN LA TRANSCRIPCIÓN DEL HPV
 - 3.9. EL FACTOR NUCLEAR CTCF
 - 3.9.1. CTCF Y LOS VIRUS DE DNA
4. PLANTEAMIENTO DEL PROBLEMA
5. PREGUNTA DE INVESTIGACIÓN
6. HIPÓTESIS
7. OBJETIVO GENERAL
 - 7.1. OBJETIVOS PARTICULARES
8. RESULTADOS
 - 8.1. EVALUACIÓN *IN SILICO* DE LA REGIÓN LARGA DE CONTROL (LCR) DEL HPV TIPO 18 PARA LA PRESENCIA DE PROBABLES SITIOS DE UNIÓN DEL FACTOR NUCLEAR CTCF
 - 8.2. EVALUACIÓN DE LA PRESENCIA DEL FACTOR NUCLEAR CTCF EN LA LCR ENDÓGENA DE LA LÍNEA CELULAR HELA.
 - 8.3. CONSTRUCCIONES PLASMÍDICAS QUE PERMITEN INSERTAR AL GEN REPORTERO EGFP BAJO EL CONTROL DE LA LCR DE HPV TIPO 18 EN LA LÍNEA CELULAR HELA

**8.4. EVALUACIÓN DE LA PRESENCIA DE CTCF EN EL TRANSGÉN ESTABLE
EN CLONAS OBTENIDAS DE LA LÍNEA CELULAR HELA.**

9. DISCUSIÓN DE RESULTADOS

10. CONCLUSIONES

11. PERSPECTIVAS

12. METODOLOGÍA

13. REFERENCIAS

1. RESUMEN

En los países en vías de desarrollo, el 80% de los cánceres que se presentan en la población son ocasionados por infecciones con agentes patógenos y la mitad de ellos por virus. Se ha demostrado que los virus poseen estrategias proliferativas muy eficientes asociadas a mecanismos anti-apoptóticos que, cuando actúan, alteran los mecanismos de homeostasis del hospedero y desencadenan cambios en las células, lo que conlleva al desarrollo de la transformación celular y del cáncer.

Dentro de estos agentes infecciosos virales se encuentran los papilomavirus humanos (HPVs), los cuales se asocian con más del 5% de los cánceres en el humano. La familia *Papillomaviridae* está constituida por virus epiteliotrópicos no envueltos, con un genoma circular de doble cadena de aproximadamente 8 kb. Aunque se han identificado más de 450 diferentes genotipos de papilomavirus humanos, solamente un pequeño grupo de ellos, denominados de alto riesgo, infectan específicamente a las células epiteliales de las mucosas y son los responsables de la mayoría de los casos de cáncer cervical y anal. La infección con el virus del papiloma humano 16 y/o 18 constituye el factor de riesgo más importante para el desarrollo del cáncer cervicouterino (CaCU). El CaCU es la segunda forma de cáncer más común de la mujer en México y en el 2018, ésta representó el 9.6% del total de muertes por neoplasias malignas en nuestro país.

Aunque se ha avanzado mucho en el conocimiento de los mecanismos de patogenicidad del virus del papiloma, poco se sabe acerca de los mecanismos epigenéticos asociados con la capacidad del virus de expresar su genoma, luego de haberse insertado establemente en el genoma de su hospedero. Se ha demostrado que los sitios de integración del virus del papiloma están aleatoriamente distribuidos en el genoma; sin embargo, se ha reportado que los llamados “sitios frágiles” del genoma, son sitios preferidos para la integración del genoma de este virus. Las regiones teloméricas están constituidas por secuencias repetidas de tipo “TTAGGG”, que se encuentran en los extremos de los cromosomas y ahí se han identificado una gran cantidad de “sitios frágiles”. La cromatina de las regiones teloméricas se encuentra en estado de heterocromatina y, se sabe que, a pesar de que estas regiones impiden la expresión génica, las secuencias repetidas que se encuentran en estas regiones se transcriben para dar origen a RNAs no codificantes que, mediante mecanismos de interferencia mediada por RNA se encuentran íntimamente relacionadas con el silenciamiento génico y la estabilidad e integridad cromosómica, experimentalmente no permiten la transcripción de los genes insertados de forma exógena.

Dada la existencia de sitios frágiles en estas zonas del cromosoma, no es de sorprender que existan reportes de integraciones de copias de HPV. Lo sorprendente es que estas copias son activas en su transcripción, por lo que debe existir un mecanismo por el cual el virus pueda mantener su capacidad para expresarse, aun cuando se encuentre en presencia de un ambiente cromatínico represivo, que de forma natural no permite la expresión de los transgenes.

Recientemente se identificó que el genoma del virus del papiloma humano posee secuencias de unión para el factor nuclear CTCF (CCCTC-binding Factor). El CTCF es una proteína conformada por 11 dedos de zinc y tiene la capacidad de unirse tanto al DNA como a otras proteínas. Se han descrito diversas funciones para esta proteína, por ejemplo, la formación de asas de cromatina al interior del núcleo, la capacidad de bloquear la actividad de los enhancers y especialmente la capacidad de contender en contra de la expansión de la heterocromatina y es precisamente esta capacidad de remodelamiento de la cromatina, la que podría estar protegiendo al virus del silenciamiento epigenético cuando se integra en regiones de heterocromatina constitutiva.

El presente proyecto estuvo enfocado en demostrar y caracterizar si existía la unión del factor nuclear CTCF con la región larga de control (LCR) del virus del papiloma humano HPV tipo 18 integrados en regiones de heterocromatina, mediante la construcción de plásmidos que contienen la secuencia de la LCR de HPV tipo 18, unidas al gen reportero de la proteína verde fluorescente (EGFP) y con esta herramienta, se evaluó la capacidad que pudieran tener los complejos proteicos celulares unidos a la LCR de contender contra la represión de la heterocromatina y si CTCF pudiera ser importante para este proceso. Las clonas obtenidas de la línea celular HeLa transfectadas con el transgén "LCR-EGFP" en regiones teloméricas demostraron un comportamiento diferencial en la actividad del transgén EGFP. Ensayos de Inmunoprecipitación de la Cromatina acoplados a q-PCR (ChIP-qPCR) indicaron que el factor nuclear "CTCF" no se une a estas regiones y por lo tanto no participa en la regulación de la expresión genética del transgén EGFP en este modelo, y que este comportamiento pudiera ser explicado a través de la presencia de las marcas epigenéticas H3K9me3 y H3K27me3 en estos transgenes. Las clonas con mayor actividad EGFP fueron aquellas que no mostraron un enriquecimiento en las marcas represivas H3K9me3 y H3K27me3 mientras que aquellas que tienen una actividad parcial o nula se enriquecen en esta marca epigenética. Estos resultados indican que deben existir mecanismos epigenéticos independientes de CTCF que regulan al HPV tipo 18 y que

dependen de H3K9me3 para poder silenciarse lo que deja abierta la posibilidad de que existan otros mecanismos moleculares que estén participando en este fenómeno.

2. ABSTRACT

In developing countries, 80% of cancers that occur in the population are caused by infections with pathogens and half of them by viruses. Viruses have been found to have very efficient proliferative strategies associated with anti-apoptotic mechanisms that, when they act, alter the host's homeostasis mechanisms and trigger changes in cells, which leads to the development of cell transformation and cancer. Among these viral infectious agents are human papilloma viruses (HPV), which are associated with more than 5% of human cancers. The Papillomaviridae family consists of non-enveloped epitheliotropic viruses, with a double-stranded circular genome of approximately 8 kb. Although more than 450 different human papillomavirus genotypes have been identified, only a small group of them, termed high-risk, specifically infect mucosal epithelial cells and are responsible for most cases of cervical and anal cancer. Infection with human papillomavirus 16 and / or 18 is the most important risk factor for the development of cervical cancer (CaCU). CaCU is the second most common form of cancer in women in Mexico and in 2018, it represented 9.6% of all deaths from malignant neoplasms in our country.

Although much progress has been made in understanding the pathogenicity mechanisms of the papilloma virus, little is known about the epigenetic mechanisms associated with the ability of the virus to express its genome, after being stably inserted into the genome of its host. Papillomavirus integration sites have been shown to be randomly distributed across the genome; however, the so-called "fragile sites" of the genome have been reported to be the preferred sites for the integration of the genome of this virus. The telomeric regions are made up of repeated sequences of the "TTAGGG" type, which are found at the ends of the chromosomes and many "fragile sites" have been identified there. The chromatin of the telomeric regions is in the heterochromatin state and, although these regions prevent gene expression, it is known that the repeated sequences found in these regions are transcribed to give rise to non-coding RNAs that through RNA-mediated interference mechanisms are closely related to gene silencing and chromosomal stability and integrity, experimentally they do not allow the transcription of exogenously inserted genes. Given the existence of fragile sites in these areas of the chromosome, it is not surprising that there are reports of HPV copy integrations. What is surprising is that these copies are active in their transcription, so there must be a mechanism by which the virus can maintain its ability to express itself, even when it is in the presence of a repressive chromatin environment, which naturally does not allow transgene expression.

The human papillomavirus genome was recently identified as having binding sequences for the nuclear factor CTCF (CCCTC-binding Factor). CTCF is a protein made up of 11 zinc fingers and can bind both DNA and other proteins. Various functions have been described for this protein, for example, the formation of chromatin loops within the nucleus, the ability to block the activity of enhancers and especially the ability to protect against the expansion of heterochromatin and it is precisely this chromatin remodeling capacity, which could be protecting the virus from epigenetic silencing when it integrates into constitutive heterochromatin regions.

The present project was focused on demonstrating and characterizing whether there was the union of the nuclear factor CTCF with the long control region (LCR) of the human papillomavirus HPV type 18 integrated in regions of heterochromatin, by means of the construction of plasmids that contain the sequence of the HPV type 18 LCR, linked to the reporter gene of the green fluorescent protein (EGFP) and with this tool, the ability of cellular protein complexes linked to the LCR to contend against heterochromatin repression and whether CTCF could be important for this process. The clones obtained from the HeLa cell line transfected with the "LCR-EGFP" transgene in telomeric regions showed a differential behavior in the activity of the EGFP transgene. qPCR coupled to Chromatin Immunoprecipitation assays (ChIP-qPCR) indicated that the nuclear factor "CTCF" does not bind to these regions and therefore does not participate in the regulation of gene expression of the EGFP transgene in this model, and that this behavior could be explained through the presence of the epigenetic marks H3K9me3 and H3K27me3 in these transgenes. The clones with the highest EGFP activity were those that did not show an enrichment in the repressive marks H3K9me3 and H3K27me3, while those with partial or no activity were enriched in these epigenetic marks. These results indicate that there must be epigenetic mechanisms independent of CTCF that regulate HPV type 18 and that depend on H3K9me3 in order to be silenced which leaves open the possibility that there are other molecular mechanisms that are participating in this phenomenon.

3. ANTECEDENTES

3.1. Cáncer Cervicouterino

El cáncer cervicouterino (CaCu) es un importante problema de salud pública, se encuentra clasificado como el tercer cáncer más frecuente en mujeres a nivel mundial (**Figura 1.**). Los datos estimados para el 2018 fueron de aproximadamente 570,000 casos nuevos por año, siendo los países en desarrollo como África, Asia, América Central y América del Sur, los lugares donde se presenta mayor tasa de incidencia con el 85% del total de los casos. ^{1,23}

El número de muertes que se siguen registrando para el CaCu es de gran relevancia, ya que en la actualidad está considerado una de las principales causas de muerte relacionadas con cáncer entre la población femenina, con más de 280,000 muertes en el mundo por año, lo cual representa más de 50% de los casos diagnosticados ^{1,4}

Estimated number of new cases in 2018, worldwide, all cancers, females, all ages

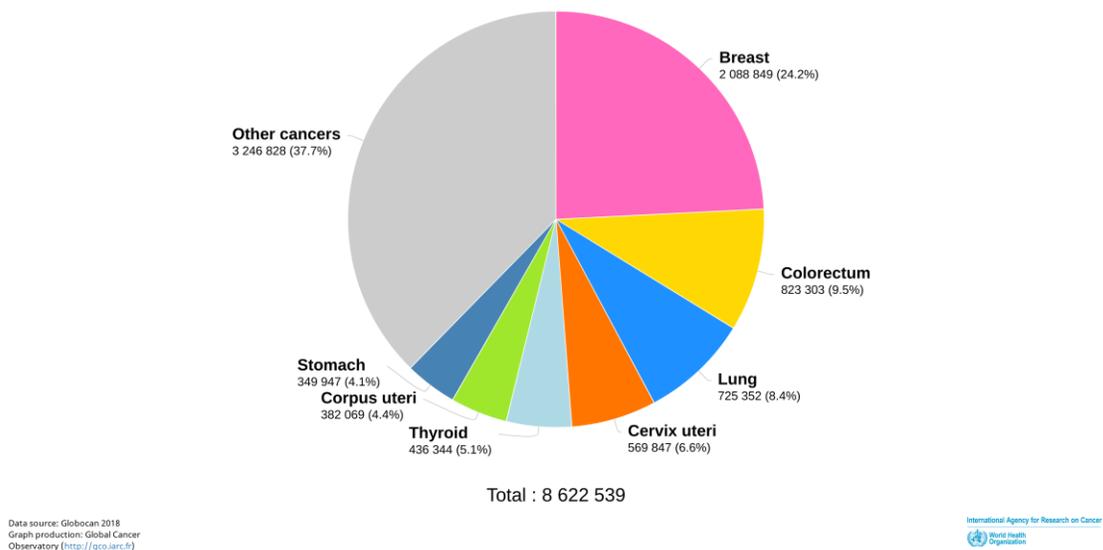


Figura 1. Incidencia de los principales tipos de cáncer en el mundo en mujeres. (GLOBOCAN 2018). El cáncer cérvicouterino se clasifica en el tercer lugar.

En lo que corresponde a nuestro país (México), históricamente el CaCu ha sido uno de los cánceres más tratables, pero a pesar de ello, sigue siendo un importante problema de salud pública en mujeres mayores de 25 años, registrándose una muerte cada 2 horas^{5,6}, con lo

cual representa la tercera causa de muerte en mujeres en las últimas décadas, con una tasa de crecimiento anual de 0.76%⁷, siendo superado únicamente por el cáncer de mama y de tiroides¹ (**Figura 2.**). Si bien se han hecho avances importantes en la detección, diagnóstico y tratamiento de esta enfermedad, desafortunadamente esto no ha cambiado drásticamente el pronóstico.⁴

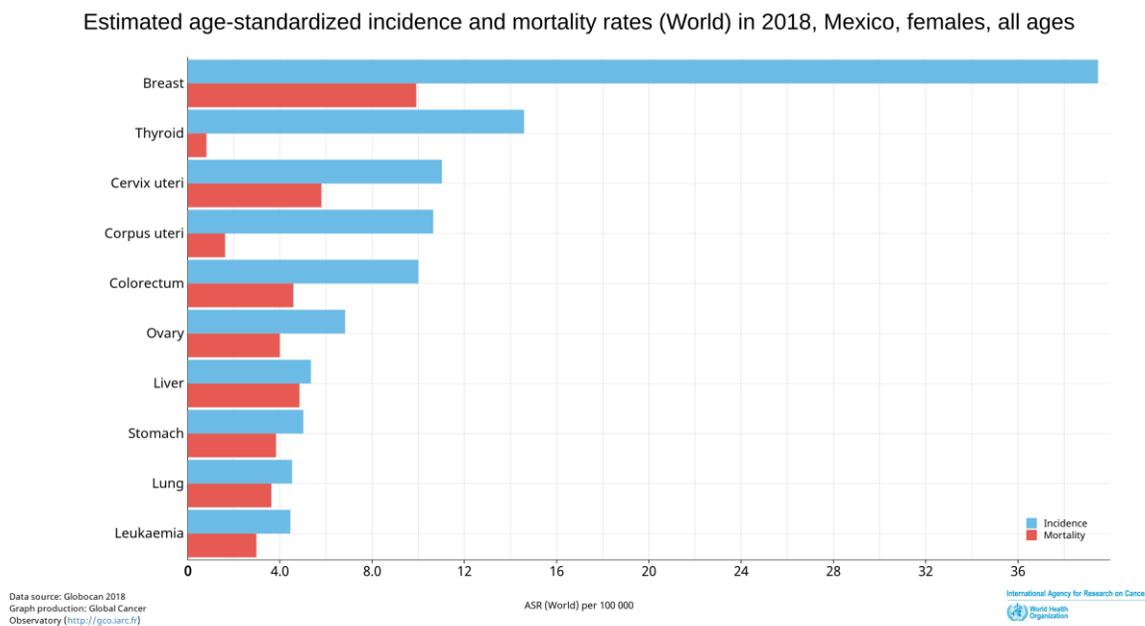


Figura 2. Incidencia y mortalidad de los principales tipos de cáncer en México. (GLOBOCAN 2018). Las barras color azul indican incidencia, barras en color rojo indican la mortalidad.

3.2. Etiología del Cáncer Cervicouterino

Durante mucho tiempo se sospechó de un agente infeccioso causante de verrugas y papilomas en humanos. En 1842, el médico italiano Rigoni-Stern asoció el desarrollo del cáncer cervicouterino con la actividad sexual.⁸ A finales de 1960 y con el auge de la microbiología comenzaron a aparecer los primeros reportes que hacían responsable a la infección viral con el virus Herpes simple tipo 2 y haciendo a este último candidato responsable de la etiología del cáncer cervical. Finalmente, estos resultados no pudieron ser confirmados. Fue hasta la década de los 70s que el grupo de Harald zur Hausen pudo demostrar, mediante experimentos de hibridación, que las verrugas genitales y los tejidos de cáncer de cérvix contenían genomas del virus del papiloma humano.⁹⁻¹¹

3.3. Papilomavirus Humanos (HPV) y Cáncer Cervicouterino

Aproximadamente el 12% de todos los cánceres humanos son ocasionados por virus, donde más del 80% de los casos se encuentran en países en vías de desarrollo.¹ A pesar de su prevalencia, importancia en la salud pública e idoneidad para el diseño de terapias dirigidas, los cánceres inducidos por virus aún enfrentan retos desafiantes. Esto debido a los limitados modelos animales de la enfermedad, la diversidad de cánceres inducidos por virus, los distintos tipos de virus que los causan y a la complicada naturaleza de las interacciones entre el hospedero y el virus, que conduce al desarrollo del cáncer.⁷ El 5% de todos los cánceres humanos en el mundo son causados por papilomavirus humanos (HPVs). Papillomaviridae, es una familia de virus epiteliotrópicos no envueltos con un genoma circular de DNA de doble cadena de aproximadamente 8 Kb. De los 450 genotipos de papilomavirus descritos, el grupo de alfa papilomavirus, denominado de alto riesgo, infectan específicamente a las células de las mucosas y son los responsables de casi todos los casos de carcinoma cervical, el causante principal de muerte por cáncer en mujeres, así como otros cánceres anogenitales, incluyendo cáncer de vulva, vagina, pene, ano y una fracción de carcinomas orales. En cambio, los virus del papiloma humano de bajo riesgo causan verrugas genitales benignas. HPV 16 y 18 son los virus de alto riesgo más abundantes, mientras que HPV 6 y 11 son los virus de bajo riesgo más abundantes en la población.⁴

En presencia de micro lesiones, los HPVs infectan a las células epiteliales basales, donde los genomas virales se mantienen persistentemente como episomas con un bajo número de copias. Los procesos de amplificación de los genomas virales, expresión de genes tardíos y la síntesis de la progenie viral están confinados a las capas terminales de diferenciación de los epitelios infectados. Debido a que las células epiteliales diferenciadas terminales se encuentran arrestadas en el ciclo celular y por lo tanto no replican su genoma, los HPVs han evolucionado para mantener a estas células en un estado favorable para la replicación del DNA. Las estrategias que los papilomavirus han desarrollado para mantener a las células del hospedero replicando su genoma tienen el potencial para desencadenar la aparición del cáncer.⁶

3.4. Estructura y organización del genoma de los Papilomavirus Humanos

Los virus del papiloma humano (HPVs) forman una gran familia de virus pequeños, no envueltos, de doble cadena de DNA, y son la principal causa de verrugas y de proliferaciones epiteliales benignas. Hasta principios de los años 70s, se pensó que solamente existía un solo tipo de HPV y que este único HPV ocasionaba lesiones verrugosas en un amplio rango de tejidos¹². En tan solo una década y con el auge de las tecnologías del DNA recombinante y de la biología molecular, fue claro que existían múltiples tipos de HPVs diferentes y que los diferentes tipos de verrugas que formaban se debían a distintos tipos de HPV, con diferentes tropismos por superficies mucosas o cutáneas.¹³ Hasta hoy en día, se conocen al menos 450 genotipos de HPVs diferentes. Estos virus no se clasifican en serotipos, sino con base en su secuencia de DNA. Los virus del papiloma humano (HPVs) tienen predilección por superficies epiteliales cutáneas o mucosas y se clasifican en dos grupos: los HPVs de bajo riesgo (que predominantemente ocasionan verrugas benignas) y los de alto riesgo que pueden ocasionar lesiones malignas¹⁴. Todos los papilomavirus codifican en su genoma para 4 proteínas altamente conservadas: Proteínas tempranas, E1 y E2 (factores de la replicación), y las proteínas de expresión tardía, L1 y L2 que son proteínas de la cápside. Además, los papilomavirus en su genoma pueden codificar proteínas accesorias: E4, E5, E6 y E7. Estas proteínas modulan el ambiente celular para favorecer la replicación viral y para evadir al sistema inmunitario.¹⁵

El genoma del HPV consiste en una molécula circular de DNA de doble cadena, de aproximadamente 8000 pares de bases (pb). Se divide en tres regiones: la región larga de control, (LCR), que no posee marco de lectura, la región que corresponde a las proteínas tempranas (E1 a E7) y la región que corresponde a las proteínas tardías (L1 y L2). **(Figura 3.)**

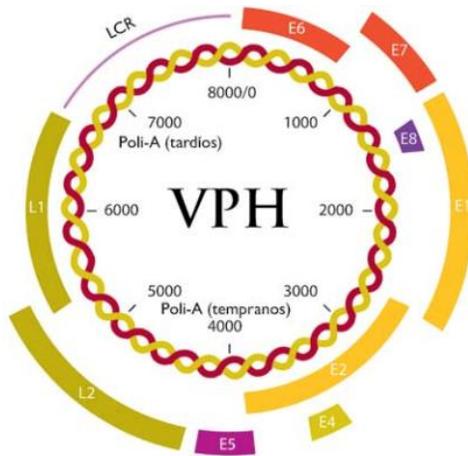


Figura 3. Genoma del HPV. Molécula circular de DNA de doble cadena, de aproximadamente 8000 pb. (Tomado de: Saavedra y Lizano, 2006).

3.5. El genoma celular, la cromatina, su estructura y su regulación

El genoma es el conjunto completo de todas las secuencias de información genética de un organismo. Éste proporciona toda la información que el organismo requiere para funcionar. En los organismos vivos, el genoma se almacena en largas moléculas de DNA llamadas cromosomas¹⁶. Pequeñas secciones de DNA, denominadas genes, codifican el RNA y las proteínas necesarias para el funcionamiento correcto de las células y el organismo. En eucariontes, el genoma de cada célula está contenido dentro de una estructura llamada núcleo. Dentro del núcleo, el genoma celular no se encuentra distribuido aleatoriamente sino que posee una compleja arquitectura que se encuentra estructurada en dominios de cromatina.¹⁷ La cromatina es un complejo formado por DNA, RNA y proteínas asociadas que permite condensar 2 metros de longitud de DNA dentro de un núcleo de tan solo unas micras de diámetro¹⁸.

Los principales componentes proteicos de la cromatina son unas proteínas llamadas histonas. Las histonas se encuentran altamente conservadas en organismos eucariontes, lo que sugiere que la estructuración en forma de cromatina evolucionó de un ancestro común de todos los eucariontes y de acuerdo con esta hipótesis, se han encontrado equivalentes de estas proteínas y estructuras simplificadas de la cromatina en organismos unicelulares del reino Archaeabacteria.¹⁹

La subunidad fundamental de la cromatina es el nucleosoma, que consiste en aproximadamente 165 pares de bases (pb) de DNA envuelto alrededor de un octámero de histonas (dos de cada histona H2A, H2B, H3 y H4). Cada nucleosoma se encuentra conectado por un segmento corto de DNA conocido como DNA conector o “*linker*” (de aproximadamente 10 a 80 pb de longitud).²⁰ Mediante imágenes obtenidas a través de espectroscopía electrónica fue posible determinar la forma nativa de la cromatina que está compuesta por una cadena de nucleosomas de aproximadamente 11 nm. Posteriormente, gracias las técnicas de microscopía de super resolución fue posible determinar que esta fibra de 11 nm forma agregados heterogéneos denominados “clutches” de aproximadamente de 30 a 50 nm. Estos “clutches” pueden unirse para formar dominios de cromatina (TADs) y su organización es específica para cada tipo celular.²¹

Nuestro entendimiento acerca de la organización cromosómica dentro del núcleo ha ido cambiando paralelamente con el poder de resolución de las técnicas moleculares e imagenológicas. La microscopía de campo claro nos permitió conocer que los cromosomas ocupan regiones específicas dentro del núcleo (territorios cromosómicos). El desarrollo de técnicas como 3C, ChIP-seq y Hi-C confirmaron la presencia de estos territorios cromosómicos, así como numerosos niveles de organización dentro del núcleo incluyendo a los compartimentos A y B, dominios asociados topológicamente (TADs) así como estructuras del genoma formadas por interacciones dentro de los TADs.²¹ **(Figura 4.)**

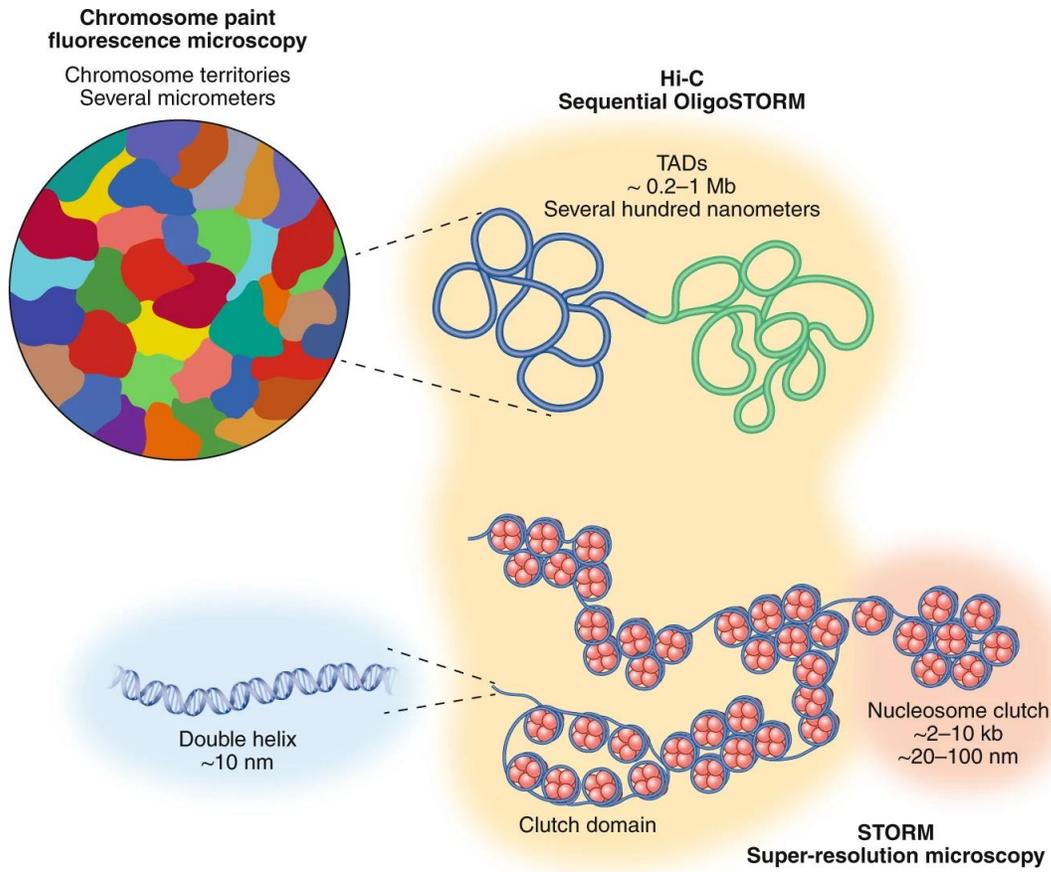


Figura 4. Dentro de la célula, los cromosomas individuales ocupan regiones específicas, conocidas como territorios cromosómicos, que miden varios micrómetros de longitud. Los enfoques genómicos y las imágenes de súper resolución revelaron la existencia de dominios que se asocian topológicamente (TADs) en el rango de decenas de kilobases a megabases, o varios cientos de nanómetros. Las imágenes de super resolución revelaron además nano dominios de nucleosomas (o “clutches”) de alrededor de unas pocas kilobases en tamaño genómico, así como la agrupación de estos nano dominios en estrecha proximidad para formar dominios más grandes con un rango de tamaño similar al de los TAD. (Tomado de: María Pía Cosma & Melike Lakadamyali, 2020).

Las señales reguladoras que ingresan al núcleo encuentran a la cromatina y no al DNA libre. La respuesta bioquímica que conduce a la activación de la expresión génica implica, por lo tanto, cambios en la estructura de la cromatina. La naturaleza compacta de la cromatina es inaccesible y, por lo tanto, es deficiente para que sucedan procesos como la transcripción, en la que el DNA de doble cadena debe servir como plantilla para RNA polimerasa²².

Los componentes de la cromatina están sujetos a una amplia gama de modificaciones que están correlacionadas con la actividad génica. Existen al menos tres formas en las que la estructura de la cromatina puede ser alterada. Primero, se puede inducir la remodelación de la cromatina a través de complejos específicos ATP dependientes. Segundo, la modificación covalente de las histonas puede ocurrir dentro de los nucleosomas²³ y tercero, las variantes de histonas pueden reemplazar una o más de las histonas que forman el “core”²⁴. Algunas modificaciones afectan la estructura o estabilidad del nucleosoma directamente, mientras que otras introducen grupos funcionales químicos que son reconocidos por proteínas reguladoras. La cromatina puede participar tanto en la activación como en la represión de la expresión de los genes y esto representa una regulación que va más allá de la regulación contenida únicamente en la secuencia de nucleótidos de dichos genes. Estos mecanismos de regulación se conocen como mecanismos regulación epigenética²⁵.

La regulación epigenética comprende a todos aquellos mecanismos que influyen en la expresión heredable de los genes sin que existan cambios en la secuencia del DNA²⁶. Existen al menos seis procesos asociados con este tipo de regulación: modificaciones postraduccionales de las histonas, metilación del DNA, complejos remodeladores de la cromatina dependientes de ATP, complejos Polycomb y Trithorax, RNAs no codificantes y la organización tridimensional de la cromatina, así como la dinámica nuclear. Estos procesos regulan la expresión génica y se encuentran íntimamente correlacionados.²⁶

Muchas de las modificaciones postraduccionales de las histonas ocurren en aquellos aminoácidos que se encuentran en las regiones amino terminales de las histonas. Estos incluyen residuos de lisinas, que pueden ser acetilados, metilados o acoplados a ubiquitina; argininas, residuos que pueden estar metilados; y serinas, residuos que pueden estar fosforilados. Todas las modificaciones pueden afectarse unas a otras, y muchas están positiva o negativamente correlacionadas entre sí. Colectivamente, este conjunto de marcadores ha sido denominado como el código de las histonas.²⁷ La modificación de las histonas es un proceso dinámico. La cromatina asociada con genes transcripcionalmente activos se enriquece en histonas acetiladas y las enzimas responsables de la acetilación o desacetilación de histonas se reclutan a sitios donde la expresión génica se activará o reprimirá, respectivamente. La metilación en ciertos residuos de aminoácidos de histonas es importante para la activación, mientras que en otros sitios es una señal de inactivación transcripcional.²² Por ejemplo, la trimetilación de la lisina 4 de la región amino-terminal de

la histona H3 (H3K4me3) es una marca epigenética que está generalmente asociada con elementos promotores de genes transcripcionalmente activos, mientras que las trimetilaciones en las lisinas 9 y 27 de la histona H3 (H3K9me3 y H3K27me3) son modificaciones típicas de regiones heterocromáticas donde no se espera actividad transcripcional.

En la cromatina se presentan dos estructuraciones principales, por un lado, la forma llamada eucromatina, está poco condensada y puede transcribirse. Este tipo de cromatina está constituida por regiones ricas en genes, pobres en secuencias repetidas, susceptibles a la digestión por enzimas como la DNasa I, presenta abundancia de la RNA Pol II y se replica tempranamente durante la fase S del ciclo celular.

La segunda forma, llamada heterocromatina, está altamente condensada y generalmente no se transcribe. Este tipo de cromatina se caracteriza por contener regiones con una baja densidad génica, es abundante en elementos repetidos, se encuentra altamente compactada y presenta abundancia de la proteína de la heterocromatina 1 (HP1) así como una replicación tardía durante la fase S del ciclo celular.²⁸

Existen dos tipos de heterocromatina: constitutiva y facultativa. Aunque ambos tipos se pueden heredar de manera estable a través numerosas divisiones celulares, cada tipo tiene una composición y una función distintos. La heterocromatina constitutiva comprende principalmente elementos genéticos repetitivos, como los telómeros y centrómeros, que se localizan en la periferia nuclear. Por el contrario, la heterocromatina facultativa es inducible y puede formarse en cualquier lugar del núcleo²⁹. Su formación es necesaria para el silenciamiento génico de los loci que regulan el apareamiento en levaduras, la inactivación del cromosoma X y la progresión del desarrollo en células de mamíferos²⁹.

Se cree que la heterocromatina constitutiva actúa principalmente como un estabilizador del genoma que previene los reordenamientos genéticos entre secuencias genéticas muy similares, así como un mecanismo de defensa en contra de DNA invasor y asegura una segregación cromosómica eficiente. Los genes adyacentes a secuencias repetidas pueden silenciarse mediante la propagación de este tipo de heterocromatina: este fenómeno se conoce como variegación por efecto de posición (PEV) que ha sido ampliamente estudiado en *Drosophila*, levaduras y en humanos, aunque en estos últimos ha sido poco estudiado, existe el efecto de posición telomérico (TPE).³⁰ El efecto de posición telomérico (TPE) provoca el silenciamiento reversible de un gen cerca de un telómero. La fuerza de este

silenciamiento depende tanto de la longitud de los telómeros como de la distancia al gen a los mismos.³¹

Los telómeros son estructuras especiales de heterocromatina que se encuentran en los extremos de los cromosomas y que son esenciales para proteger estas regiones de la recombinación y de la degradación del DNA³². En los vertebrados, los telómeros están compuestos por repetidos en tándem de la secuencia TTAGGG junto con proteínas específicas que se unen a estas secuencias. Son el sustrato de la enzima telomerasa, una DNA polimerasa que tiene la capacidad de alargar estas secuencias.³² La mayoría de los tejidos somáticos no poseen actividad telomerasa y se presenta acortamiento telomérico con cada división consecutiva. Sin embargo, ciertamente se ha observado la actividad regulada de la telomerasa en células basales de tejidos altamente regenerativos, como los del sistema inmune, piel, mucosas e intestino³³. La actividad de la telomerasa puede ser detectada *in vitro* desde la fase G1, pero los telómeros se replican *in vitro* hasta la fase S tardía del ciclo celular. Durante la mayor parte del ciclo celular, la telomerasa se acumula en los cuerpos de Cajal, y en la fase S se traslada y se une a los telómeros en la célula. Esto significa que existe una regulación a nivel de la localización espacial de telomerasa activa y el telómero (sustrato)³⁴.

Los telómeros de los mamíferos, además de ser sustratos para la telomerasa y para factores de unión a secuencias repetidas, también están unidos por nucleosomas.

Estos nucleosomas contienen variantes de histonas con modificaciones características de la heterocromatina constitutiva. La variante de histonas H3.3 se expresa en todas las etapas del ciclo celular y puede ser incorporada a la cromatina independientemente de la replicación del DNA. Además de ser una marca de genes activos, la histona H3.3 es depositada por el complejo de chaperonas ATRX/DAXX en regiones heterocromáticas, que incluye telómeros y heterocromatina pericéntrica³⁵. La deposición de la histona H3.3 es esencial para la colocación de la marca epigenética H3K9me3 requerida para mantener reprimida a la cromatina en los telómeros³⁵.

Los telómeros también se encuentran enriquecidos en la proteína de la heterocromatina 1 (HP1), poseen altos niveles de histona H3.3 trimetilada en la lisina 9 (H3K9me3), histona H3.3 trimetilada en la lisina 27 (H3K27me3) y la histona H4 trimetilada en la lisina 20 (H4K20me3).³⁶ **(Figura 5.)**

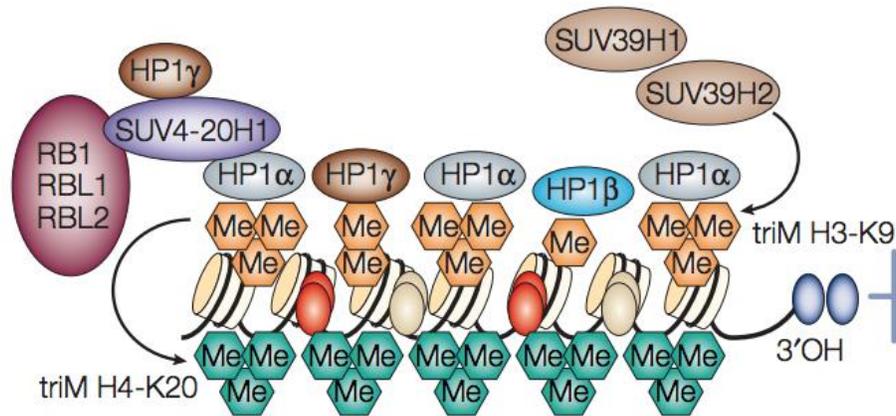


Figura 5. Regulación epigenética de los telómeros. Los telómeros están unidos por nucleosomas que contienen histonas que llevan a cabo modificaciones que son características de la heterocromatina constitutiva. El supresor de variegación (SUV39H) es una metiltransferasa de histonas que juega un papel importante en la trimetilación de la histona H3 en el residuo de lisina 9 (H3K9me3). Esta marca a su vez recluta eficazmente a las isoformas de la proteína 1 de la heterocromatina (HP1) (HP1 α , HP1 β y HP1 γ) a estas regiones. HP1 recluta al supresor de variegación (SUV4-20H), que a su vez trimetila a la histona H4 en el residuo de lisina 20 (H4K20me3). (Modificado de: Blasco., 2005.)

3.6. Integración del genoma del Virus del Papiloma humano al genoma celular

La infección persistente por un subconjunto de virus de papiloma humano (HPV) denominado de "alto riesgo" puede promover el desarrollo del cáncer. En estos cánceres, el genoma viral extra cromosómico se encuentra integrado en el genoma del huésped.³⁷ Mientras que la oncogenicidad de los HPVs es consecuencia de las estrategias replicativas de estos virus, los cánceres asociados a HPVs generalmente representan a infecciones no productivas donde algunas proteínas virales se expresan, pero no se produce ninguna progenie viral infecciosa. Todos los carcinomas positivos para HPVs de alto riesgo contienen secuencias integradas de HPV en el genoma celular³⁷. Los papilomavirus tienen un ciclo de vida que aprovecha el proceso natural de renovación del tejido en epitelios estratificados. Los virus del papiloma explotan este proceso; acceden e infectan las células basales a través de micro abrasiones y establecen una infección a largo plazo en estas células en división. Los HPVs se encuentran integrados en lesiones premalignas y en una gran variedad de cánceres anogenitales y orofaríngeos^{38,39}, pero esto no forma parte del

ciclo de vida viral. A diferencia de la integración sitio específica de los retrovirus, que es catalizada por una integrasa viral y es un paso obligatorio en su proceso de replicación, la integración del DNA del HPV en los cromosomas del huésped es un punto muerto del ciclo de vida viral y es considerado un evento aleatorio inespecífico, aunque se han propuesto como blancos de integración a los sitios frágiles del DNA del huésped, regiones intergénicas y sitios activos transcripcionales⁴⁰. Al ser integrado, se pierde parte del genoma viral y éste a su vez pierde la capacidad de producir genomas virales circulares, formar viriones y transmitirse a un nuevo huésped³⁷. La integración parcial del HPV resulta en una expresión desregulada de los oncogenes virales E6 y E7, lo que promueve la proliferación celular, anula los puntos de control del ciclo celular y promueve la inestabilidad genómica³⁷. Esto le confiere a las células una ventaja de crecimiento y promueve la progresión oncogénica³⁷.

Casi todos los eventos de integración del HPV que se han estudiado con detalle hasta la fecha están relacionados con la oncogénesis provocada por este virus. Los eventos de integración del HPV se pueden detectar en lesiones premalignas, pero el porcentaje de células que contienen HPV integrado aumenta a medida que las células progresan hacia el cáncer⁴¹. La integración parcial del genoma del HPV generalmente da como resultado una desregulación de la expresión de los oncogenes virales E6 y E7, que promueve la proliferación celular, anula los puntos de control del ciclo celular y provoca inestabilidad genómica. Esto le da a las células una ventaja de crecimiento y promueve la progresión oncogénica⁴².

A pesar de que se han realizado numerosos esfuerzos por encontrar sitios específicos de integración del HPV, solamente se han logrado identificar algunos loci recurrentes. Lo que sí es claro, es que estos "hotspots" genómicos están altamente correlacionados con sitios frágiles⁴³ y regiones transcripcionalmente activas del genoma⁴⁴. El 50% de los cánceres cervicales positivos para el HPV16 y el 65% de los positivos para el HPV18 contienen al genoma de HPV integrado en alguno de estos sitios frágiles.⁴⁵ Los sitios frágiles son loci cromosómicos específicos altamente inestables que, bajo condiciones específicas de cultivo o exposición a ciertos agentes químicos, tienden a formar rupturas cromosómicas y otras anomalías cromosómicas en respuesta a estrés en el proceso de replicación del DNA⁴⁶.

Los sitios frágiles comunes son regiones cromosómicas particularmente sensibles a la formación de rupturas en los cromosomas metafásicos después de la inhibición parcial de la síntesis de DNA⁴⁷. Existen aproximadamente 90 sitios frágiles comunes (CFS) descritos

en todo el genoma humano⁴⁸. Estas lesiones se observaron por primera vez en células cultivadas en condiciones de estrés inducido por la ausencia de folato para inducir la expresión citogenética del sitio X frágil (FRAXA) en individuos con retraso mental ligado al cromosoma X frágil⁴⁷. Más tarde se descubrió que estos sitios frágiles también eran inducidos específicamente por la afidicolina, un inhibidor de las DNA polimerasas α , δ y ϵ ⁴⁷ y fueron posteriormente caracterizados a través de técnicas citogenéticas en cromosomas en metafase⁴⁹.

Los sitios frágiles comunes normalmente son estables en cultivos de células humanas⁴⁷. Sin embargo, cuando las células se cultivan en condiciones de estrés de folato o con dosis bajas de afidicolina que solo inhiben parcialmente la síntesis de DNA, estos sitios frágiles se vuelven evidentes y forman rupturas cromosómicas⁴⁷. Otros inhibidores de la replicación, como la hidroxurea, son menos específicos para inducir lesiones principalmente en sitios frágiles, probablemente debido a diferencias en los mecanismos de inhibición⁴⁷.

En el cáncer, más del 50% de las integraciones de HPV ocurren dentro de algún sitio frágil⁵⁰. Se ha demostrado que el genoma del HPV, la proteína viral E2 y la proteína celular BRD4 interactúan, forman un complejo y se asocian preferentemente en regiones frágiles del genoma humano⁵¹. Las rupturas del DNA que ocurren en los sitios frágiles activan la respuesta celular de daño al DNA (DDR)⁵². El genoma de HPV recluta a los factores (DDR) celulares y estos a su vez cargan a la DNA polimerasa celular en el genoma de HPV y se sintetizan los genomas de la progenie viral⁵². Esta asociación le permite al virus amplificar su genoma y favorece la formación de focos de replicación viral en estos sitios². Sin embargo, esto también puede resultar en una mayor integración accidental del DNA viral, que en el caso de los HPVs de alto riesgo promueve la carcinogénesis⁵¹. Esto podría explicar el porqué de la integración preferencial en estas regiones⁴⁶. De forma muy interesante para los objetivos de este trabajo, en el 2009, el grupo de De Lange y colaboradores, descubrió que los telómeros de los cromosomas poseían muchas similitudes con estos llamados sitios frágiles. Los telómeros son series muy largas de repetidos de tipo TTAGGG que se repiten muchas veces en los extremos de los cromosomas de los mamíferos. Estos repetidos se mantienen mediante la replicación semiconservativa del DNA gracias a la telomerasa. De Lange y colaboradores revelaron que las repeticiones teloméricas representan un desafío para la maquinaria de replicación del DNA, dando lugar a rupturas cromosómicas que se asemejan a las de los sitios frágiles comunes inducidos por Afidicolina.⁵³

Con respecto a la integración de DNA exógeno en sitios frágiles comunes, es importante mencionar que los telómeros frecuentemente se encuentran enriquecidos en elementos genéticos móviles. Por ejemplo, recientemente se demostró que el virus del herpes humano (HHV-6) se integra preferentemente en los telómeros de los cromosomas y que elementos LINE-1 pueden transponerse hacia los telómeros en algunas líneas celulares derivadas de hámster⁵⁴. Lo mismo parece ser cierto para el HPV debido a que existen múltiples reportes que demuestran la existencia de genomas integrados de HPV tipo 18 en múltiples sitios frágiles. Basados en la ubicación genómica de estos sitios, existe la probabilidad de que existan copias de HPV integradas en los telómeros de los cromosomas humanos.⁵⁵⁻⁵⁷

El estrés en la replicación celular de las regiones teloméricas podría tener un valor adaptativo ya que proporcionaría un lugar aparentemente seguro para la integración y posterior inactivación de elementos móviles y DNA exógeno que de otro modo podrían invadir partes más valiosas del genoma⁵³. Este efecto de posición quedó demostrado mediante la integración dirigida de transgenes hacia estas regiones. En 2007, Rincón-Arango y cols. demostraron que al dirigir la integración de transgenes hacia los telómeros de los cromosomas estos eran permanentemente inactivados. Únicamente aquellos transgenes con secuencias de tipo “insulator”, es decir, secuencias que separan o aíslan dominios funcional y estructuralmente distintos y que son capaces de reclutar a remodeladores epigenéticos, eran capaces de contender en contra de la heterocromatina asociada a este tipo de regiones y permanecer activas.⁵⁸

La integración viral del HPV al genoma de la célula resulta siempre en la pérdida de genes virales de tal forma que al integrarse casi siempre se interrumpen los genes E1 y L1/L2 perdiéndose así los genes virales E1, E2, E4, E5 pero conservando la región larga de control (LCR), región que contiene al promotor viral, así como la región que codifica para las oncoproteínas virales E6 y E7⁵⁹. Actualmente no se sabe de alguna endonucleasa celular que corte el genoma de HPV en estos sitios preferentes y se piensa que estos eventos de integración del virus están conducidos por eventos de reparación del DNA mediados por micro homología⁶⁰. De manera importante, mencionar que E2 es un represor transcripcional que regula negativamente la expresión de los oncogenes E6 y E7, ocasionando así una expresión desregulada de estos oncogenes y promoviendo la oncogénesis.⁴. Esto es de especial importancia dado que la LCR es una región no codificante del genoma del HPV que regula la transcripción de los genes virales y a la que

pueden unirse una gran variedad de factores de transcripción celulares que finalmente deciden el destino de la célula hospedera infectada.

3.7. Elementos de regulación transcripcionales del Virus del Papiloma Humano.

Los elementos de regulación del HPV modulan la transcripción de todos los genes de los papilomavirus. Estos elementos son de enorme importancia para la biología del virus, puesto que estos determinan el rango de tipos celulares que pueden infectar, acoplan la expresión de los genes tempranos (E, del inglés: early) y tardíos (L, del inglés: late) al estado de diferenciación de los epitelios, modulan la respuesta viral a hormonas y factores de crecimiento e integran y acoplan la expresión y replicación viral.

La LCR (o región larga de control) de los papilomavirus es una región genómica que no contiene ningún marco de lectura abierto y abarca aproximadamente el 10% del genoma de todos los papilomavirus. Esta región se encuentra localizada generalmente entre los genes virales E6 y L1.¹⁵

La LCR se divide en 3 regiones: la región 3', la región central y la región 5' (**Figura 6.**) La región 3' de la LCR contiene al promotor del gen temprano "E6", así como una caja TATA río arriba del sitio de inicio de la transcripción de este gen. En el caso de HPV16, su LCR también posee secuencias de unión para la proteína SP1, río arriba de la caja TATA.⁶¹ La región 3' contiene al origen de la replicación del genoma viral así como múltiples sitios de unión para las proteínas tempranas E1 y E2 que regulan y estimulan la replicación viral de forma sinérgica.⁶² La región central de la LCR de HPV 16 y HPV 18 contiene una región tipo enhancer que fue descubierta al ser clonada en vectores con genes reporteros bajo el control de promotores homólogos y heterólogos. La actividad de esta región tipo enhancer es del tipo célula específica y es muy intensa en células epiteliales pero débil en otros tipos celulares, como por ejemplo, fibroblastos.^{63,64} En efecto, los HPVs son estrictamente epiteliotrópicos y esta selectividad está en gran parte definida por la región tipo enhancer de la LCR de HPV⁶⁵. Esta región posee sitios de unión a factores transcripcionales tanto ubicuos (por ejemplo, Oct-1, NF1, AP-1) como específicos de queratinocitos (KRF-1)⁶⁶. Esta especificidad está dada por las concentraciones específicas de los factores transcripcionales ubicuos en células epiteliales pero al mismo tiempo también debe existir la unión de factores transcripcionales específicos de células epiteliales (como por ejemplo,

KRF1) para que el enhancer pueda potenciar la actividad transcripcional del genoma de HPV⁶⁶. Estos factores celulares específicos podrían representar nuevos blancos terapéuticos y ser un mecanismo interesante a desactivar para evitar el desarrollo del cáncer⁶⁷.

Contigua a la región tipo enhancer, la LCR posee una región con una actividad transcripcional mucho menor, de tipo silencer. Esta región se encuentra entre la región tipo enhancer y el promotor de la proteína E6. Como ya se mencionó anteriormente, muchos factores celulares y virales, interaccionan con esta región: por ejemplo, las proteínas virales E1 y E2, modulan la actividad transcripcional y la replicación del virus; componentes de la maquinaria basal de transcripción como SP1 y TBP; receptores de glucocorticoides que modulan positivamente la transcripción, así como otros factores que parecen intervenir en la especificidad por el tejido como KRF, Skn – 1a/i, TEF y AP1⁶⁸, es decir, que todos los virus pueden utilizar factores transcripcionales propios, pero también celulares para completar con éxito su ciclo de vida. El control transcripcional de los virus de DNA por factores transcripcionales celulares es muy importante en la regulación espacial y temporal de la expresión génica de los virus. Muchos de estos factores son reclutados para mejorar la expresión génica de los virus y, por lo tanto, aumentan la producción de la progenie viral, pero los factores de la célula huésped también pueden restringir la expresión génica del virus y la productividad de la infección⁶⁹.

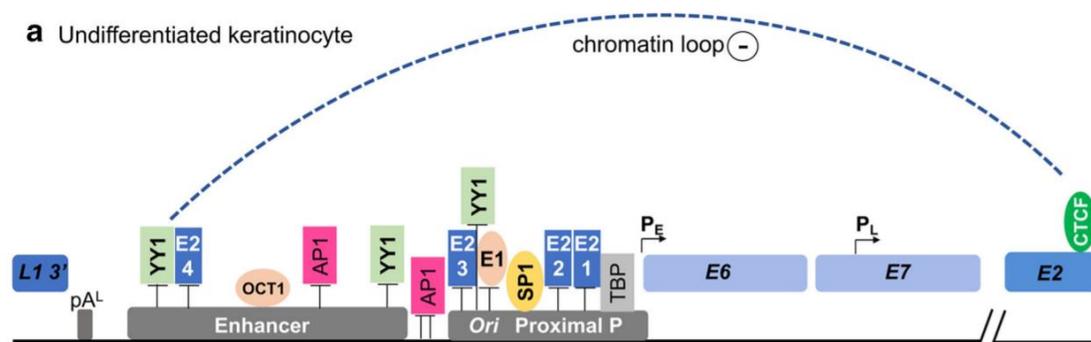


Figura 6. La LCR contiene numerosos elementos cis que controlan la expresión génica y la replicación del virus del papiloma humano (Tomado de: Burley y Parish, 2020).

3.8. Elementos genéticos que regulan la transcripción del genoma celular

La transcripción es un proceso complejo que depende de la acción coordinada de factores de unión específicos, la maquinaria basal de la transcripción, factores remodeladores de la cromatina y enzimas responsables de modificaciones covalentes de las histonas, entre otras.⁷⁰

En eucariontes, existen miles de genes codificantes y para cada uno se requiere un fino programa de control transcripcional. Mucho de este control específico depende o se ve afectado por proteínas que son capaces de unirse a secuencias específicas del DNA y que se unen a regiones reguladoras proximales y distales. A estas proteínas se les conoce como factores transcripcionales.⁷⁰

Los elementos genéticos que regulan la transcripción en eucariontes a nivel de su secuencia son: promotores, *enhancers*, *silencers* e *insulators*. Estos elementos se clasifican de acuerdo con la localización de su secuencia con respecto al sitio de inicio de la transcripción en elementos proximales o distales.⁷¹

Los promotores son aquellos loci que están sobrelapados con los sitios de inicio de la transcripción (TSS) de los genes y es en donde toda la regulación ejercida sobre un gen es integrada para obtener una tasa transcripcional. La función inmediata de un promotor es la de unir y posicionar correctamente al complejo de inicio de la transcripción, cuya actividad catalítica depende principalmente de la actividad de la RNA polimerasa dependiente de DNA.⁷² La arquitectura de los promotores comprende a los elementos proximales así como a los elementos del “core” del promotor. **(Figura 7.)** Estos elementos comprenden a los sitios de unión para factores transcripcionales, el sitio de unión a TBP, elementos tipo BRE, DRE, elementos iniciadores, entre otros.⁷²

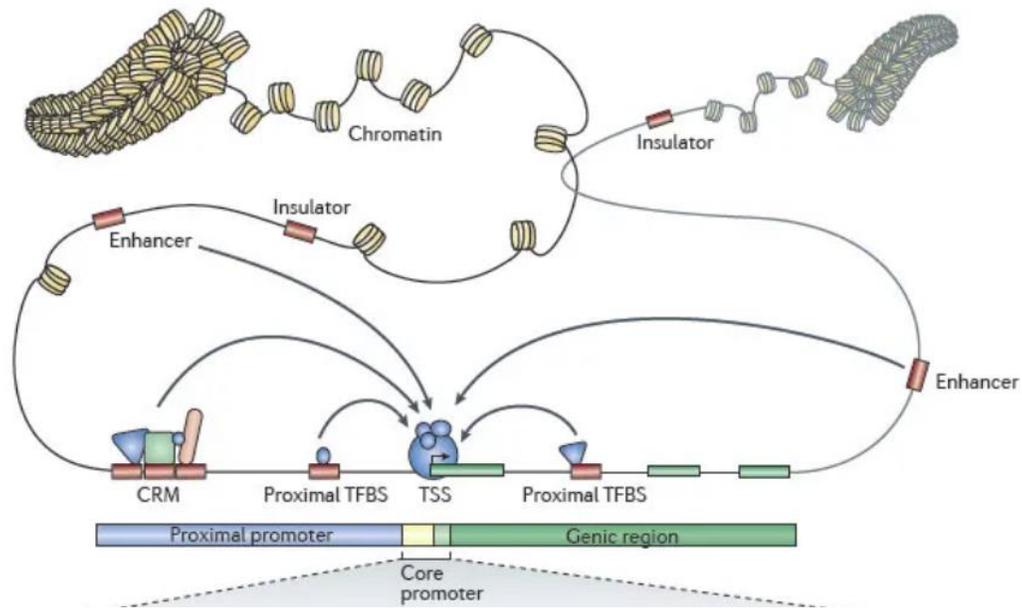


Figura 7. Elementos de regulación genéticos que regulan la transcripción de los genes. (Tomada de: Sandelin et al., 2012).

Los elementos tipo *enhancer* son aquellos elementos genéticos tejido y tiempo específicos que quedan definidos de acuerdo con su capacidad de activar la transcripción independientemente de su localización, posición u orientación con respecto a un promotor dentro de una construcción plasmídica. Esta flexibilidad es característica de los elementos tipo *enhancer* y sigue siendo parte de su definición funcional. Se encuentran frecuentemente dentro de los intrones de los genes que regulan o dentro de los intrones de los genes vecinos y, a menudo, a grandes distancias del promotor en regiones intergénicas.⁷³

Los elementos tipo *silencer* son funcionalmente opuestos a los elementos tipo *enhancer*. Estos elementos previenen la expresión génica durante la diferenciación y la progresión a través del ciclo celular.⁷⁴

Los elementos de tipo barrera o “*insulators*” son elementos muy interesantes que se encargan de aislar dominios funcionalmente autónomos que necesitan estar protegidos de la influencia de dominios vecinos y que son capaces de detener la propagación de la heterocromatina. Estas propiedades particulares son de especial interés para este trabajo.

Este tipo de elementos se pueden clasificar de acuerdo con la función que tienen como bloqueadores de *enhancer* (cuando el *insulator* se encuentra entre un promotor y un

potenciador) y barreras (cuando se ubican entre un promotor y un silenciador). La mutación o eliminación de *insulators* altera el patrón de la expresión génica y conduce a defectos de desarrollo.⁷⁵

La proteína más abundante asociada con la actividad de *insulator* es el factor nuclear CTCF. El ejemplo mejor estudiado es el del locus Igf2-H19 donde CTCF previene la activación del alelo materno Igf2 a través del bloqueo de la actividad un *enhancer* distal. Cuando CTCF se pierde, el gen se reactiva⁷⁶. Este es un ejemplo clásico, sin embargo, ahora se conocen muchas otras funciones de este factor⁷⁷. Por ejemplo, se sabe que CTCF regula los eventos de splicing de RNA mensajero, al promover la pausa de la RNA Pol II⁷⁷. Además de su unión al DNA, también se ha demostrado que CTCF interactúa directamente con el RNA⁷⁷. Un estudio basado en ensayos de CLIP-seq identificaron aproximadamente 15000 RNAs que interactúan con CTCF en mESCs⁷⁷.

3.9. El factor nuclear CTCF

El factor de unión a CCCTC (CTCF) es una proteína de unión al DNA expresada de manera ubicua y que está altamente conservada en metazoarios⁷⁸. Inicialmente descrito como un represor transcripcional del gen *Myc*⁷⁹, este factor contiene 11 dedos de zinc, 4 de los cuales (dedos de zinc 4–7) se unen fuertemente a su secuencia consenso de aproximadamente 12 pares de bases y que es común en la mayoría de los sitios de unión a CTCF⁸⁰. Los dedos de zinc restantes pueden unirse diferencialmente a otras secuencias, lo que permite un alto grado de flexibilidad y extensión del sitio de unión canónico⁸¹. El número exacto de sitios de unión a CTCF dentro del genoma humano no está claro, pero los estudios sugieren que existen hasta 26,000 sitios de unión en células humanas⁸². Estudios más recientes han destacado el potencial para muchos más sitios de unión a CTCF dentro del genoma humano (de aproximadamente 300,000 sitios) cuya ocupación depende del tipo celular y del estado de diferenciación. La unión de CTCF está enriquecida dentro de regiones intergénicas o en regiones intrónicas de los genes. Aproximadamente el 12% de la unión de CTCF se encuentra dentro de elementos proximales⁸³ y su asociación con regiones específicas del genoma está relacionada con la formación de asas de cromatina, la organización de los nucleosomas y el control de la RNA polimerasa II (RNA Pol II) para coordinar eventos de splicing de genes⁸⁴.

Los cromosomas de los mamíferos están organizados en dominios topológicos (TAD) del orden de megabases⁸⁵. Tal organización espacial es una propiedad general del genoma, es estable en diferentes tipos de células y está altamente conservado entre ratones y humanos⁸⁵. Se han identificado múltiples factores que están asociados con las regiones que limitan a los dominios topológicos, entre ellos destaca el factor nuclear CTCF, la proteína arquitectónica mejor caracterizada y la principal proteína de tipo insulador en mamíferos^{85,86}. El abatimiento de CTCF provoca una pérdida dramática del aislamiento de los TADs además de promover amplios contactos ectópicos⁸⁷. Una propiedad muy interesante de los dominios topológicos es que el par de motivos CTCF presentes en las fronteras de estos dominios se produce en una orientación convergente en más del 90% de los casos⁸⁸. La importancia de la orientación de estos motivos entre loci separados es sorprendente y debe influir en el mecanismo por el cual CTCF y la cohesina forman asas, lo que parece implicar la dimerización de CTCF⁸⁸. Experimentos en los que se altera la presencia u orientación de los sitios de unión a CTCF posibilitan la ingeniería de asas, dominios y otras estructuras de la cromatina⁸⁸.

El modelo de extrusión sugiere que las proteínas encargadas del mantenimiento estructural de los cromosomas (SMC), como la cohesina o condensina, pueden extruir progresivamente a la cromatina hasta ser bloqueadas por CTCF unido a un sitio orientado adecuadamente⁸⁹. Este modelo sugiere que los anillos de cohesina y condensina pueden atrapar topológicamente el DNA y moverse a lo largo él hasta encontrar un obstáculo que bloquea este movimiento⁹⁰. Experimentos *in vitro* han demostrado que la cohesina puede difundir a lo largo de DNA anclado y que este proceso eso puede ser bloqueado por CTCF⁹¹. Simulaciones de este modelo han sugerido que la cohesina comienza al azar en el genoma y que el proceso de extrusión es continuo hasta ser bloqueado por desde adentro del loop y se ha demostrado que la cohesina está enriquecida en el extremo 3' de los motivos de CTCF⁹². Además, la eliminación de cohesinas da como resultado la pérdida de todas las asas y dominios⁹³. Esto sugiere un mecanismo de formación de asas de cromatina, de modo que el tamaño de los dominios corresponde a la cantidad de tiempo que la cohesina es capaz de extruir cromatina antes de encontrar un obstáculo que detiene la extrusión⁸⁹.

3.9.1 CTCF Y Virus de DNA

Varios virus de DNA, incluido el virus del herpes asociado al sarcoma de Kaposi (KSHV), el virus de Epstein-Barr (VEB) y el virus del papiloma humano (HPV) son capaces de utilizar al factor nuclear CTCF para controlar su expresión génica. Muchos estudios han descrito un papel importante para CTCF en la persistencia de diversos virus oncogénicos. CTCF puede aumentar o reprimir la expresión génica de los virus y en algunos casos, CTCF puede también aumentar la complejidad de los transcritos virales obtenidos por eventos de procesamiento alternativo.⁸⁴ La infección por HPV se establece en las células basales indiferenciadas, en las que se inicia un programa estrictamente controlado por la expresión génica viral. En el contexto episomal del genoma del HPV18, Joanna Parish y sus colegas demostraron que la expresión de genes virales tempranos (E6 y E7) es atenuada en estas células indiferenciadas debido a la formación de un asa de cromatina dependiente de las proteínas CTCF y YY1⁹⁴ y no entre dos sitios convergentes de CTCF⁹⁵. A medida que las células comienzan su programa de diferenciación, se requiere de la expresión de las oncoproteínas E6 y E7 para mantener a las células en un estado proliferativo⁹⁴. Esto está coordinado por una reducción en la expresión de la proteína YY1 en tejidos con un mayor grado de diferenciación, que libera el asa de cromatina represiva y, debido a que esta proteína tiene la capacidad de reclutar a la maquinaria del complejo represivo Polycomb (PRC), ocasiona una pérdida de la marca epigenética represiva H3K27me3⁹⁶. También reduce la abundancia del factor nuclear CTCF en el ORF del gen viral E2 necesario para la formación del asa⁹⁴. El aumento en la accesibilidad y la reducción de H3K27me3 en el enhancer viral resultan en la activación del promotor temprano de E6, un aumento en el reclutamiento de la RNA Pol II y en la deposición de la marca epigenética activadora H3K4me3, lo que resulta en un aumento de la expresión de los oncogenes E6 y E7 ⁹⁴.

(Figura 8.)

La mutación del único sitio de unión E2-CTCF en HPV18 no tiene un efecto en la replicación o en el mantenimiento de los episomas de HPV18, pero resulta en la producción temprana de transcritos virales y en un aumento en la expresión de las oncoproteínas E6 y E7, así como hiperproliferación celular⁹⁵. De manera interesante, los datos de ChIP-seq disponibles en ENCODE para las células HeLa indica que si bien el sitio de unión a CTCF dentro del ORF E2 se mantiene en el DNA integrado de HPV18, el factor nuclear CTCF no se detecta en este sitio⁹⁵. Esto puede indicar que la represión de la transcripción viral a través de la organización de la cromatina CTCF-YY1 se anula en las células cancerosas⁹⁵. Debido a

que los sitios de unión a CTCF se encuentran frecuentemente mutados en cáncer, la pérdida de unión de CTCF en el DNA viral integrado en los tumores puede representar un evento de progresión en el desarrollo del cáncer⁹⁷.

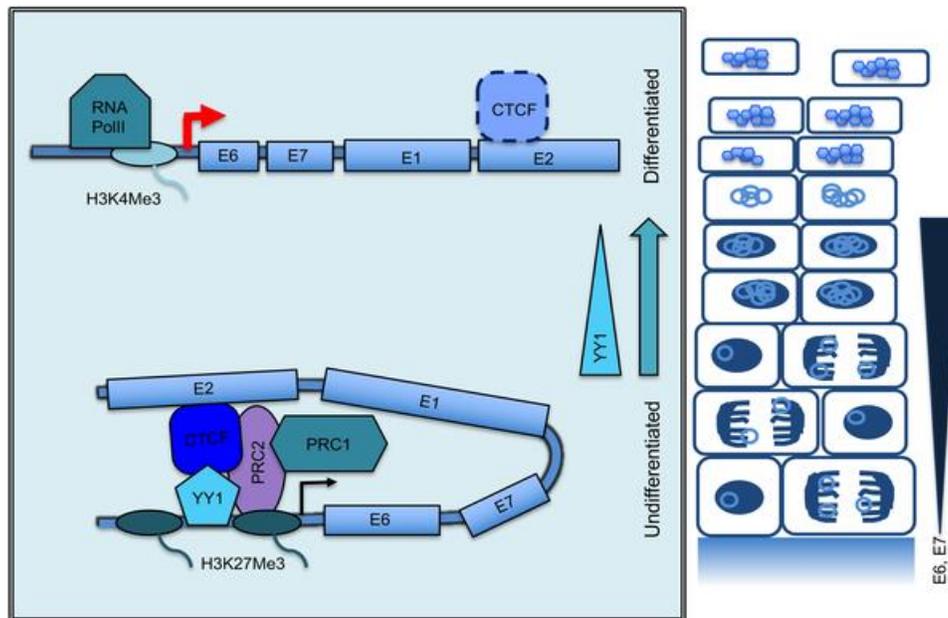


Figura 8. Modelo del control epigenético de la transcripción de los oncogenes de HPV dependiente de diferenciación. (Tomada de: Parish et al., 2018).

Finalmente es importante mencionar que poco o nada se sabe acerca de las funciones que pudieran tener los sitios de unión a CTCF dentro de la LCR una vez que el genoma del HPV ha sido integrado al genoma de la célula. Episomalmente, al perder el ORF del gen viral E2, CTCF no puede unirse a la LCR del virus. Este trabajo busca entender si una vez integrado existe o se recupera esta interacción y de ser así, si ésta pudiera proteger al genoma viral en contra del silenciamiento inducido por la maquinaria celular dependiendo de la posición genómica donde ha sido integrado el virus.

4. PLANTEAMIENTO DEL PROBLEMA

Aunque se ha reportado que los HPVs pueden integrarse en regiones teloméricas, poco se sabe acerca de los mecanismos epigenéticos responsables del mantenimiento de la actividad promotora de la LCR de los HPVs después de su integración en estas regiones. Estudiar los mecanismos epigenéticos de protección de transgenes virales y su relación con el mantenimiento de su actividad transcripcional contribuirá al entendimiento molecular de los mecanismos protectores de los virus en contra del silenciamiento epigenético inducido por la maquinaria celular. Esto conllevaría repercusiones clínicas importantes debido a que la integración del HPV en el genoma de la célula es un paso clave en el desarrollo del cáncer que resulta en la sobre expresión de los oncogenes virales E6 y E7 lo que brindaría la posibilidad de proponer nuevas estrategias terapéuticas contra la carcinogénesis inducida por la infección con HPV.

5. PREGUNTA DE INVESTIGACIÓN

Investigar si la LCR del HPV 18 conserva su actividad transcripcional después de ser integrada en una región de heterocromatina constitutiva. Y de ser así, caracterizar la participación de CTCF en el mantenimiento de su actividad en estas regiones.

6. HIPÓTESIS

Existen mecanismos mediados por factores epigenéticos independientes de CTCF que participan para prevenir el silenciamiento de la región larga de control (LCR) del virus del papiloma humano (HPV) tipo 18 después de ser integrada en regiones de heterocromatina telomérica en células HeLa.

7. OBJETIVO GENERAL

Caracterizar si existen mecanismos epigenéticos mediados por CTCF implicados en la regulación transcripcional de la LCR de HPV tipo 18, cuando se integra en una región de heterocromatina telomérica.

7.1. OBJETIVOS PARTICULARES

1. Evaluar *in silico* los posibles sitios de unión de CTCF en la región de la LCR del HPV18.
2. Comprobar que CTCF se une *in vitro* a la LCR de HPV tipo 18 en líneas celulares derivadas de cáncer cervicouterino.
3. Generar las construcciones plasmídicas que permitan insertar a la LCR del HPV 18 y al gen reportero EGFP en regiones de heterocromatina telomérica en la línea celular HeLa derivada de un adenocarcinoma positivo a HPV tipo 18.
4. Obtener y caracterizar clonas transfectadas establemente con estas construcciones en células HeLa.

8. RESULTADOS

8.1. Evaluación *in silico* de la región larga de control (LCR) del HPV tipo 18 para la presencia de probables sitios de unión del factor nuclear CTCF

Como primer paso, se realizó un estudio *in silico* empleando el programa (CTCFBSDB 2.0) que es una herramienta diseñada para facilitar los estudios sobre elementos tipo *silencer* y sus funciones genómicas. Este programa contiene una colección de aproximadamente 15 millones de sitios de unión a CTCF (CTCFBS) determinados experimentalmente⁹⁸. Para este propósito se analizaron las secuencias de la región 5', la región central, así como la región 3' de la LCR de HPV tipo 18 respectivamente. De acuerdo con las predicciones sugeridas por este programa, la región larga de control del HPV tipo 18 contiene 2 sitios de unión para CTCF con los scores más altos dentro de la región 5' de la LCR, así como en la región promotora proximal respectivamente (Figura 9). Estos sitios son similares en secuencia a aquellos reportados por (Schmidt et al., 2012) y que constituyen a dos de los sitios de unión a CTCF más conservados en mamíferos.⁹⁹ **(Figura 9. Esta figura fue realizada con los datos del análisis *in silico* indicando las regiones de la LCR de HPV18.)**

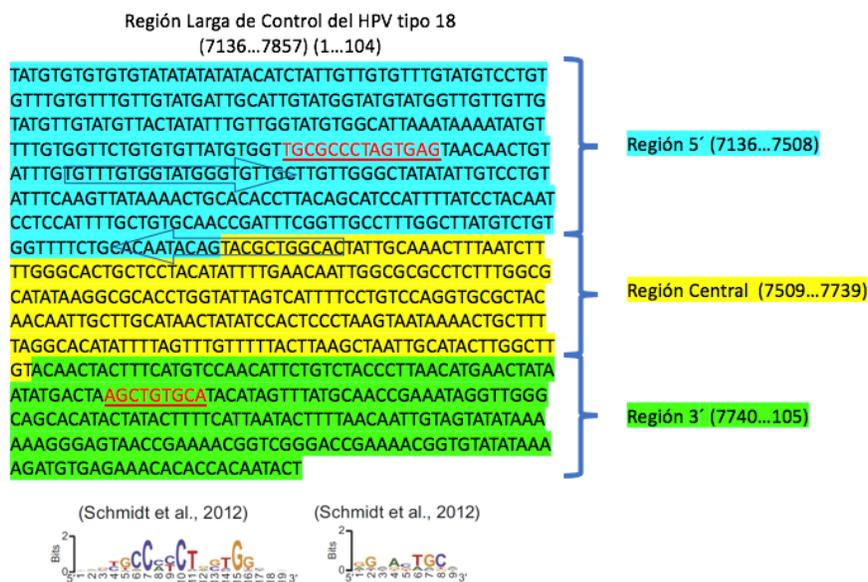


Figura 9. Representación gráfica de la región larga de control (LCR) del HPV tipo 18 mostrando los posibles sitios de unión para el factor nuclear CTCF dentro de la región 5' y la región 3' de la LCR. En rojo se muestran los sitios identificados por el programa CTCFBSDB 2.0 que

guardan similitud con los reportados por Schmidt et al., 2012. Las flechas indican los oligonucleótidos empleados para la evaluación de la LCR endógena por qPCR.

8.2. Evaluación de la presencia del factor nuclear CTCF en la LCR endógena de la línea celular HeLa.

Para determinar la presencia del factor nuclear CTCF en la región larga de control (LCR) endógena del HPV tipo 18 se realizaron ensayos de ChIP-qPCR en la línea celular HeLa, que es una línea celular derivada de adenocarcinoma de cérvix positiva a la infección por HPV tipo 18 y mantiene la expresión de los oncogenes E6 y E7 ^{100,101}. Estos ensayos revelaron que en la línea celular HeLa, CTCF se encuentra presente en al menos una LCR de todas las copias integradas del virus HPV tipo 18. Se observó un enriquecimiento de dos veces por encima del control negativo para el anticuerpo anti-CTCF, el exón 27 del gen Rb, y por encima del anticuerpo irrelevante IgG.

Para este ensayo se empleó un par de oligonucleótidos situados entre los dos sitios de unión probables para la unión de CTCF en las regiones 5' y 3' de la LCR de HPV tipo 18. En las **(Figuras 9 y 10)** se observó un valor de enriquecimiento del factor nuclear CTCF dos veces mayor en la LCR endógena de HPV tipo 18 de la línea celular HeLa que en el exón 27 del gen Rb (control negativo de la técnica de ChIP) y que en el anticuerpo irrelevante IgG, mientras que el control positivo (gen WRAP 53) obtuvo un valor de más de 500 veces por encima del control negativo y del anticuerpo irrelevante IgG. Realmente el valor obtenido en la LCR no es suficiente para decir que CTCF se está uniendo en esta región ni tampoco quiere decir que se esté uniendo en todas las 50 copias de la LCR en esta línea celular. Además, no se sabe cuántas de estas LCR se encuentran en regiones teloméricas ni si la unión de CTCF ocurre en estas LCR particulares. A pesar de que la diferencia entre la LCR y el control negativo (Exón 27 Rb) es significativa, las limitaciones de la técnica y de los oligonucleótidos diseñados no permiten diferenciar las integraciones de las LCR y realmente son un promedio de las más de 50 copias de esta secuencia integradas en todo el genoma de esta línea celular por lo que fue necesario diseñar un modelo que permitiera integrar esta secuencia en regiones teloméricas para poder medir su actividad sin la contribución de las LCR endógenas sin perder los factores transcripcionales necesarios que permitan su correcta actividad.

Análisis de la ocupación de CTCF en la LCR de HPV18 en la línea celular HeLa

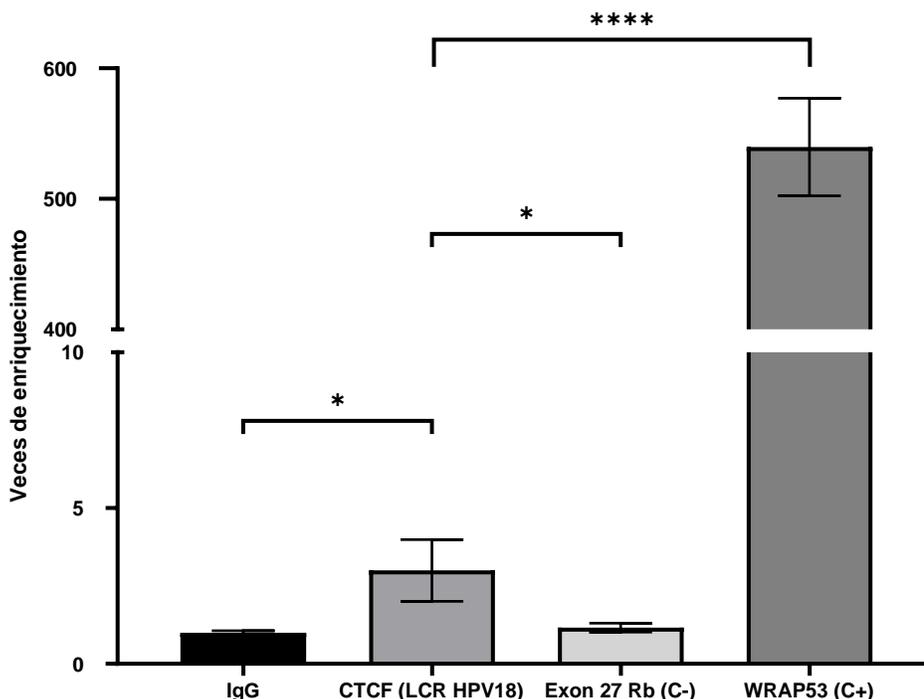


Figura 10. CTCF se une débilmente en las más de 50 copias LCR endógena del HPV 18 *in vitro* en células HeLa. Ensayo de ChIP-qPCR para determinar la presencia de CTCF evaluando la LCR HPV tipo 18 endógena de la línea celular HeLa. Se presenta el enriquecimiento contra IgG. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). El anticuerpo inespecífico IgG se empleó como control negativo de la técnica de ChIP. El análisis de PCR cuantitativa (qPCR) se realizó con DNA obtenido de la técnica de ChIP que evalúa la región comprendida entre los dos sitios probables de unión a CTCF de la LCR de HPV 18 (CTCF LCR HPV18) y dos regiones control para el enriquecimiento de CTCF. Como control negativo se empleó al exón 27 del gen Rb (C-) y como control positivo al gen WRAP 53 (C+). (Guerra-Calderas, 2018).

De esta manera, se observó el enriquecimiento promedio de CTCF en las copias endógenas de la LCR en la línea celular HeLa. Esto hizo necesario el diseño de un modelo celular que permitiera el estudio de copias de la LCR integradas únicamente en regiones teloméricas.

8.3. Construcciones plasmídicas que permiten insertar al gen reportero EGFP bajo el control de la LCR de HPV tipo 18 en la línea celular HeLa.

Con la finalidad de crear un sistema que permitiera la evaluación de la LCR de HPV tipo 18 integrada en regiones teloméricas en la línea celular HeLa sin la influencia de las más de 50 copias de HPV que posee esta línea celular, se construyeron plásmidos que contuvieran a la LCR de HPV tipo 18 dirigiendo la expresión del gen reportero EGFP. Este gen reportero se vuelve muy útil para ser estudiado por ensayos de citometría de flujo a través del tiempo. Para este fin, se procedió a clonar la LCR de HPV tipo 18 mediante PCR empleando como templado al plásmido PGL2-LCR18 (que es un plásmido que contiene a la LCR de HPV tipo 18 acoplado al gen reportero de la Luciferasa)¹⁰² y un par de oligonucleótidos sintéticos que alinean en los extremos de la LCR junto con dos extremos de corte para la endonucleasa BamHI. Una vez que se obtuvo el producto amplificado de la LCR de HPV tipo 18, se procedió a clonar este fragmento en el vector pJET (Thermo, K1231) para su posterior verificación y digestión con la enzima de restricción BamHI. Se observó que las clonas 12, 15, 16, 20, 01, 05 y 10 contenían al inserto LCR de HPV tipo 18 (**Figura 11.**). El inserto LCR fue posteriormente purificado y ligado al plásmido pEGFP1. Esta nueva construcción fue analizada mediante PCR con un par de oligonucleótidos que alinean en la LCR y en el gen reportero EGFP para verificar la orientación de la LCR. Solamente las clonas 2, 6, 8, 9 y 10 contenían a la LCR en la orientación deseada y permisiva a la transcripción del gen reportero EGFP (**Figura 12.**). Este plásmido fue posteriormente verificado por secuenciación Sanger. Para la posterior introducción del fragmento telomérico se digirió el plásmido pSTY11 (donado por la Dra. De Lange del Instituto Rockefeller)¹⁰³, que es un plásmido que contiene 800 pb del repetido telomérico TTAGGG, con las enzimas de restricción PstI y BglII para su posterior ligación en el plásmido pEGFP1-LCR18 anteriormente construido. Como se muestra en la (**Figura 12 y 13**), al digerir con la enzima BamHI, todas las construcciones liberan un fragmento de aproximadamente 800 bp que corresponde con el tamaño esperado para la LCR de HPV tipo 18. Únicamente el plásmido que contiene al repetido telomérico (pEGFP1-LCR18-Telo) libera un fragmento de 800 bp al ser digerido con las endonucleasas BglII y PstI.

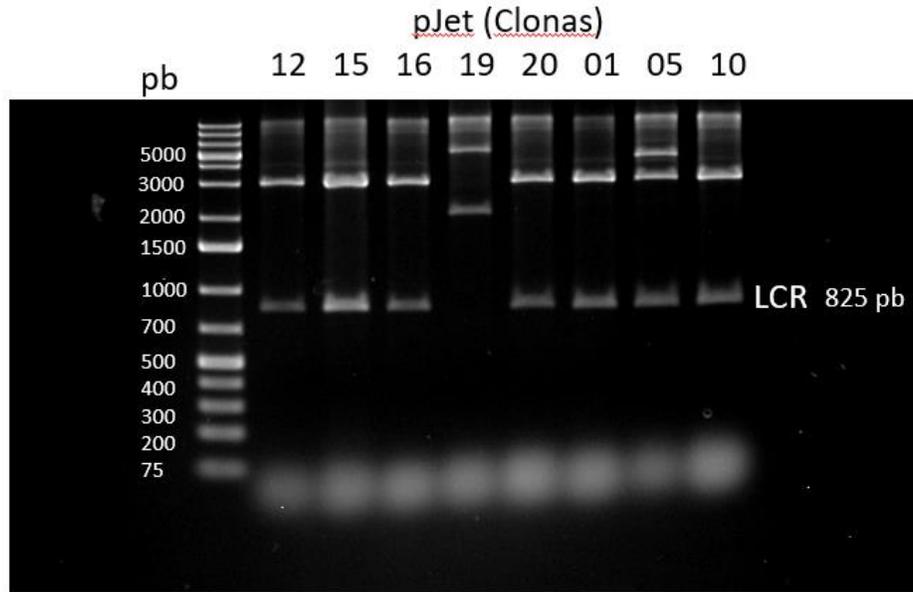


Figura 11. Digestión del plásmido vector pJET con la endonucleasa de restricción BamHI. Se observa un fragmento liberado de aproximadamente 800 pb que corresponde con el tamaño esperado para la LCR.

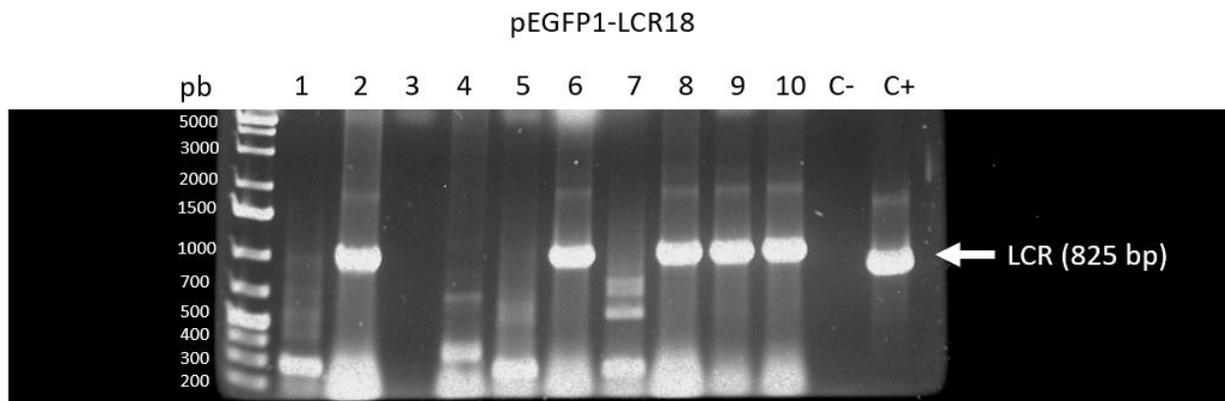


Figura 12. PCR de Clonas que poseen al plásmido pEGFP1-LCR18 en la orientación correcta de la LCR de HPV tipo 18 que permite la transcripción del transgén EGFP. Únicamente las clonas 2, 6, 8, 9 y 10 muestran el fragmento esperado de 825 bp.

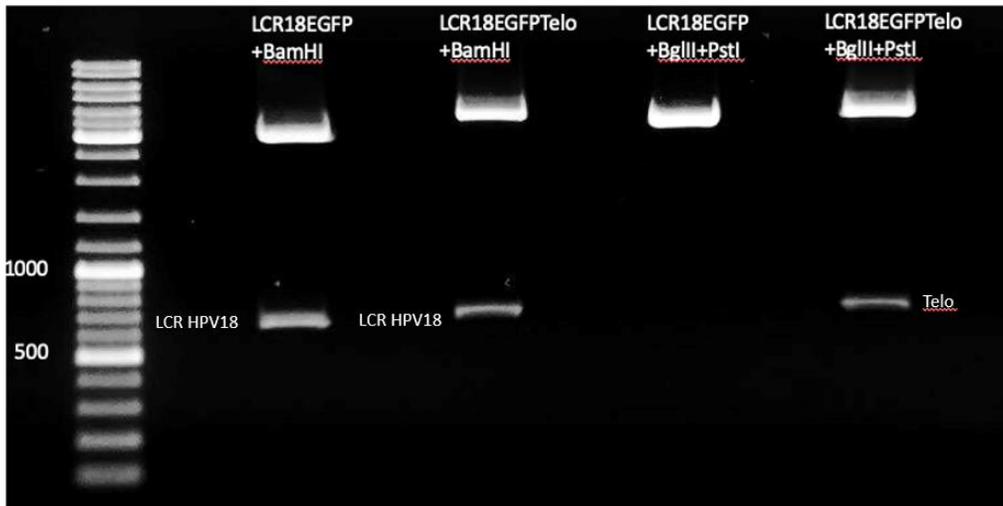
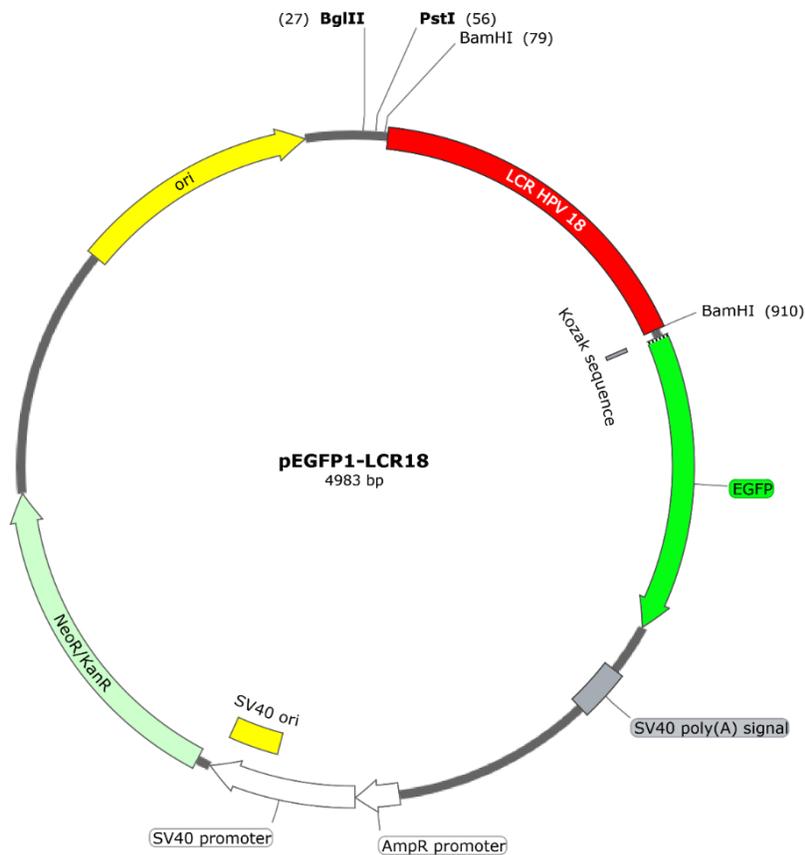


Figura 13. Corrimiento electroforético de los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo digeridos con las endonucleasas de restricción BglIII y PstI. Se libera un fragmento de 800 pb cuando se digiere con la enzima BamHI que corresponde con el tamaño esperado para la LCR y otro fragmento de 800 pb al digerir con las enzimas BglIII y PstI que corresponde con el tamaño esperado para el fragmento telomérico.



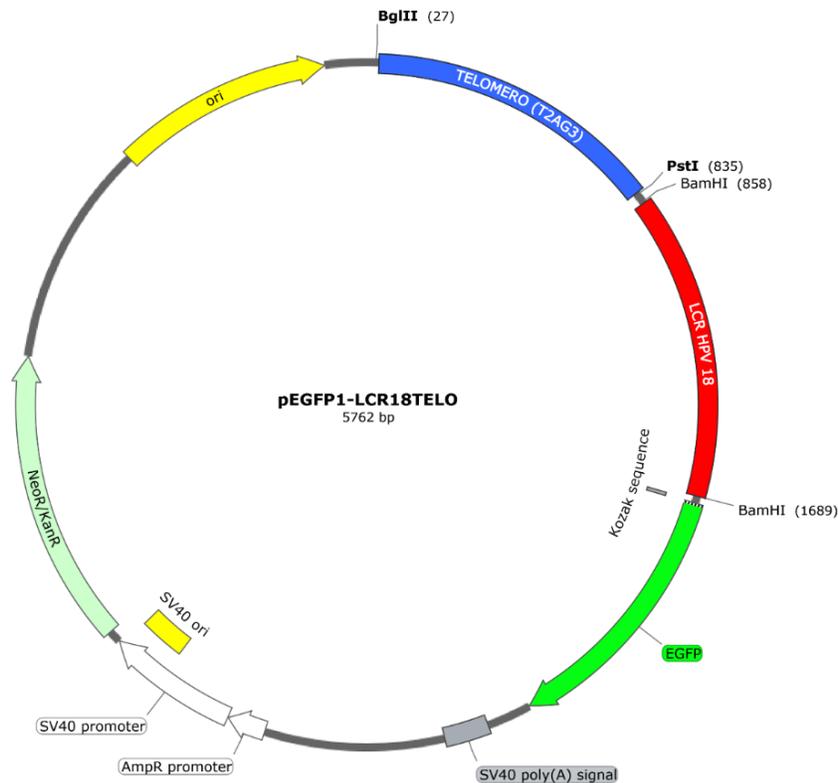


Figura 14. Representación gráfica de los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo. Los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo, de tamaño 4983 y 5762 pb respectivamente, contienen a la LCR de HPV tipo 18 controlando la expresión del gen EGFP. Particularmente el plásmido pEGFP1-LCR18-Telo contiene un fragmento telomérico que favorecerá la integración en los telómeros de células HeLa, además de contar con los genes que confieren la resistencia a los antibióticos ampicilina, kanamicina y neomicina respectivamente.

De esta manera se logró construir plásmidos que contienen a la región larga de control (LCR) del HPV tipo 18 controlando la expresión del gen EGFP y un fragmento telomérico que permitirá favorecer la integración de los transgenes hacia las regiones teloméricas en células HeLa para poder evaluar la actividad transcripcional de la LCR integrada en estas regiones. Además, la construcción contiene los genes que confieren a las células transfectadas la resistencia al antibiótico neomicina y esto nos permitirá seleccionar clonas transfectadas establemente con estas construcciones para poder seguir la actividad de EGFP a lo largo de 100 días postransfección.

Debido a que ya se ha visto que la actividad de la LCR del HPV depende del tipo celular que infecta, era importante emplear una línea celular donde se supiera que el virus posee actividad transcripcional. Este es el caso de la línea celular HeLa, una línea derivada de adenocarcinoma de cérvix e infectada específicamente por el HPV tipo 18. Esta línea celular contiene aproximadamente de 20 a 50 copias del HPV tipo 18. Este alto número de copias podría afectar el estudio de aquellas copias que se integran en regiones de heterocromatina. Por esta razón se realizó la construcción de plásmidos que expresaran al gen reportero EGFP bajo el control de la LCR de HPV tipo 18.

Estos plásmidos, pEGFP1-LCR18 y pEGFP1-LCR18-Telo, fueron transfectados en la línea celular HeLa con el reactivo Xfect (de acuerdo con las especificaciones del fabricante, Takara #631317) y posteriormente estas células fueron seleccionadas con el antibiótico G418 (Geneticina) a una concentración de 3 mg/mL. Se ha reportado que la presencia de 800 bp del repetido telomérico TTAGGG en el plásmido pSTY11 es suficiente para dirigir al transgén hacia regiones teloméricas, permitiendo su inserción en los telómeros de los cromosomas⁵⁸.

Después de 48 h de la transfección con estos plásmidos, las células fueron observadas en el microscopio de fluorescencia (**Figura 14**) y después de un mes de estar bajo la selección del antibiótico G418, las células fueron analizadas mediante citometría de flujo. Como se muestra en la (**Figura 15.**), en ambos casos existen dos poblaciones, una positiva a EGFP y una negativa. En aquellas células transfectadas con el plásmido sin el fragmento telomérico, existe una subpoblación del 44% que es positiva a la actividad de EGFP. Esto contrasta con las células transfectadas con el plásmido que contiene al repetido telomérico puesto que la población positiva a la actividad de EGFP es menor y es del 16.9 %. Este resultado se realizó por triplicado en dichas poblaciones celulares, sin embargo, esto también podría ser resultado de la variabilidad entre las líneas seleccionadas.

Con la finalidad de obtener poblaciones homogéneas y poder medir la actividad del transgén EGFP a través del tiempo, se procedió a obtener clonas de las poblaciones positivas y negativas transfectadas con pEGFP1-LCR18 o con pEGFP1-LCR18-Telo. Las clonas fueron obtenidas por medio del método de dilución seriada y una vez obtenidas las clonas se verificó por PCR que el transgén se encontrara integrado en el genoma celular. Para este propósito se emplearon oligonucleótidos que alinearan en el origen de la replicación del plásmido transfectado, así como en la secuencia codificante de EGFP para no amplificar copias endógenas de HPV tipo 18. Como se observa en las (**Figuras 17 y 18**),

al emplear como templado al genoma de las clonas seleccionadas, aquellas transfectadas con el plásmido pEGFP1-LCR18 amplifican una banda de aproximadamente 800 pb y aquellas que poseen el fragmento telomérico de 800 pb, amplifican una banda de aproximadamente 1500 pb.

Al mismo tiempo, se realizaron ensayos de citometría de flujo en las clonas obtenidas y se analizó su actividad durante 100 días. Como se observa en la **(Figura 19)**, de las clonas que fueron transfectadas con el plásmido pEGFP1-LCR18 (sin fragmento telomérico), la clona 001 y 006 mantuvieron la expresión de EGFP al día 100 mientras que las clonas 002 y 005 comenzaron a perder la actividad de EGFP al día 28.

De las clonas transfectadas con el fragmento telomérico, las clonas 001 y 004 pierden la expresión de EGFP al día 100, mientras que las clonas 008 y 011 mantienen parcialmente la expresión de EGFP y se observan dos poblaciones una positiva y otra negativa a la expresión de EGFP. Debido a que existe una población positiva a EGFP dentro de clona 008, se obtuvieron subclonas de esta población mediante una segunda ronda de diluciones seriadas. Como se muestra en la **(Figura 20)** se obtuvieron clonas con 3 comportamientos distintos. La subclona 8.1 es completamente positiva a la expresión de EGFP; la subclona 8.2 se divide en una población positiva y otra negativa a la expresión de EGFP y la subclona 8.3 es completamente negativa a la expresión de este transgén con respecto a la fluorescencia obtenida con células HeLa sin transfectar.

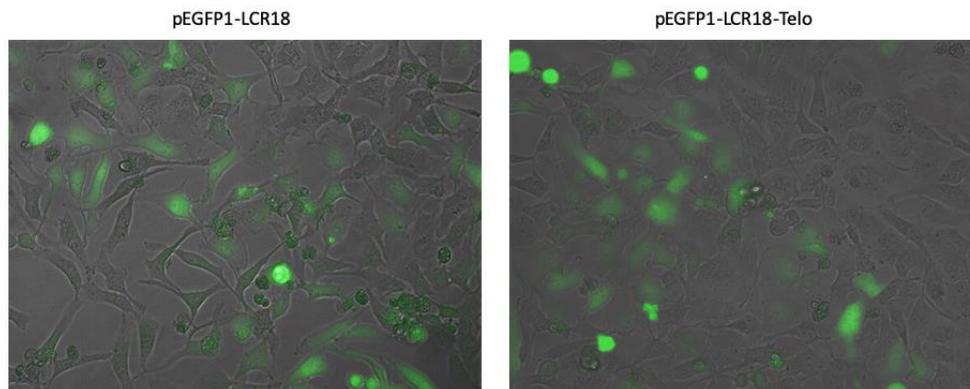


Figura. 15. Microscopía de fluorescencia de los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo (Aumento 40X) que muestra células HeLa positivas a la actividad de EGFP 48 h después de ser seleccionadas con el antibiótico G418. Se observan células al 80% de confluencia y a la proteína EGFP en el interior de las células, así como en células en división.

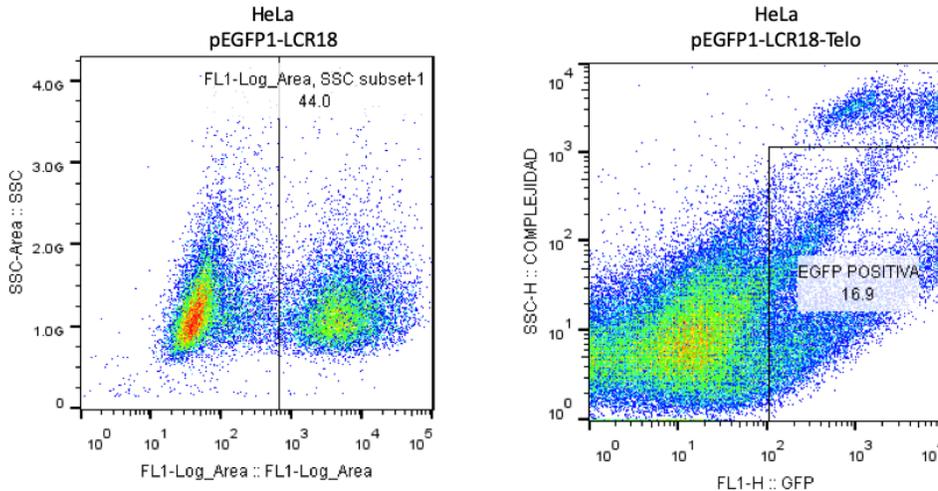


Figura. 16. Ensayos de citometría de flujo en la población total de células transfectadas con los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo respectivamente. Los ensayos de citometría fueron realizados por triplicado 30 días después de la selección con el antibiótico geneticina (G418) (incluida en los plásmidos) a una concentración de 3 (mg/mL). En el caso de las células transfectadas con el plásmido pEGFP1-LCR18 se observó una población positiva a la expresión de EGFP (44%) y una negativa (56%) mientras que en las células transfectadas con el plásmido pEGFP1-LCR18-Telo la población positiva a la expresión de EGFP fue (17%) y la negativa de (83%).

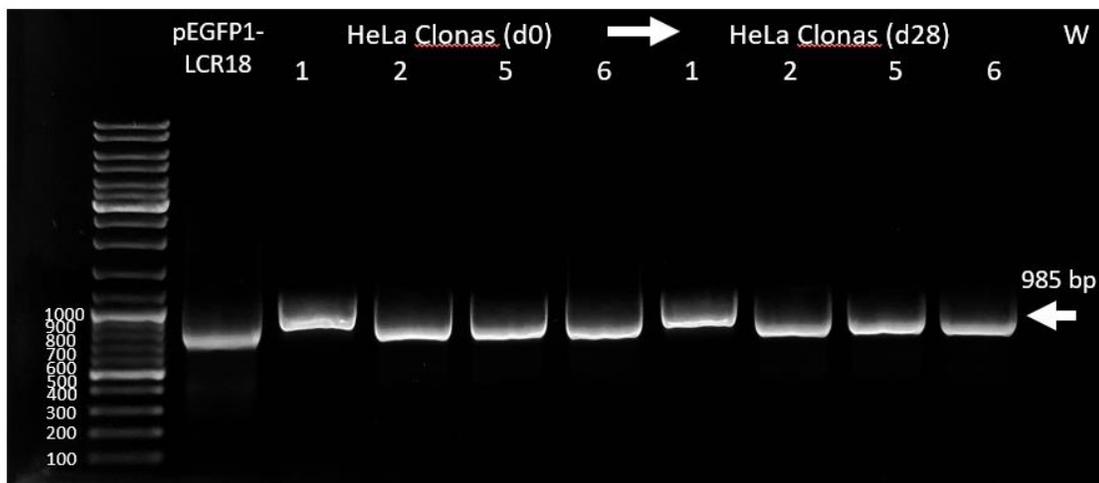


Figura 17. Ensayo de PCR punto final en clonas de la línea celular HeLa transfectadas establemente con el plásmido pEGFP1-LCR18 utilizando un par de oligonucleótidos que alinean en el origen de la replicación del plásmido y sobre el gen de la proteína verde fluorescente (EGFP). Se observó una banda de 985 pb que corresponde con el tamaño esperado para la LCR en el genoma de las clonas transfectadas con el plásmido pEGFP1-LCR18 y seleccionadas con el antibiótico G418 al día 0 y al día 28 de la evaluación de la actividad de EGFP mediante citometría de flujo.

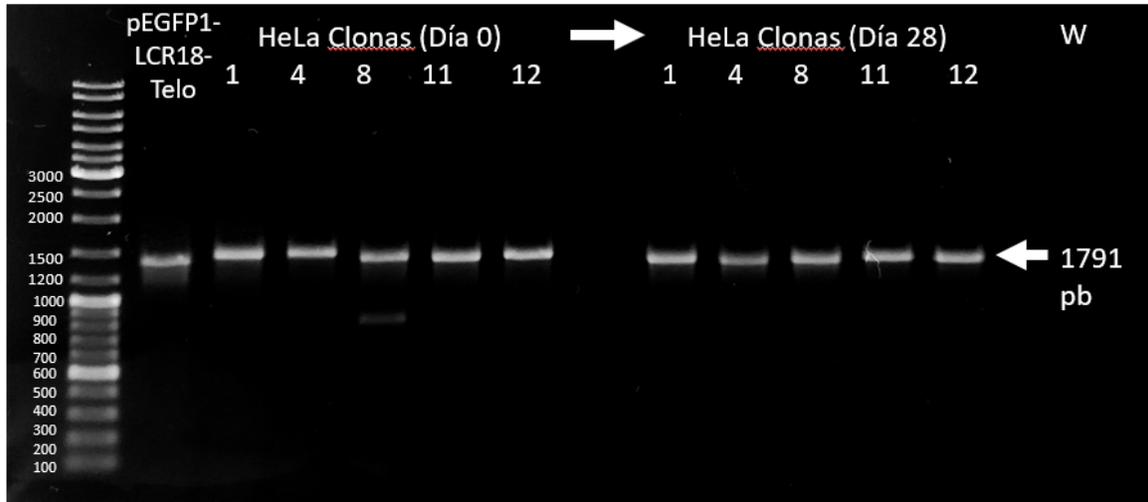


Figura 18. Ensayo de PCR punto final en clonas de la línea celular HeLa transfectadas establemente con el plásmido pEGFP1-LCR18-Telo utilizando oligonucleótidos que alinean en el origen de la replicación del plásmido y sobre el gen de la proteína verde fluorescente (EGFP). Se observó una banda de 1791 pb que corresponde con el tamaño esperado de la LCR y el fragmento telomérico en el genoma de las clonas transfectadas con el plásmido pEGFP1-LCR18-Telo y seleccionadas con el antibiótico G418 al día 0 y al día 28 de la evaluación de la actividad de EGFP mediante citometría de flujo.

Estos resultados indican que se logró la transfección estable de la línea celular HeLa con los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo y sugieren que las células transfectadas con el plásmido que contiene al fragmento telomérico (pEGFP1-LCR18-Telo) tienden a silenciarse más que aquellas transfectadas con el plásmido sin el fragmento telomérico (pEGFP1-LCR18) y esto se traduce en un enriquecimiento en la población negativa a la expresión de EGFP. Esto puede deberse a que el transgén que contiene el repetido telomérico está siendo dirigido hacia los telómeros de los cromosomas o bien pudiera tratarse de un silenciamiento mediado por efecto de posición cromosomal inducido por el fragmento telomérico de 800 pb introducido en el transgén. Estos resultados también indican que en este sistema solamente es posible evaluar la actividad de la LCR en aquellas células positivas a la actividad de EGFP. Por esta razón fue necesario obtener clonas de las poblaciones positivas y negativas a la expresión de EGFP (transfectadas con pEGFP1-LCR18 o pEGFP1-LCR18-Telo) para poder evaluar su actividad a lo largo de 100 días.

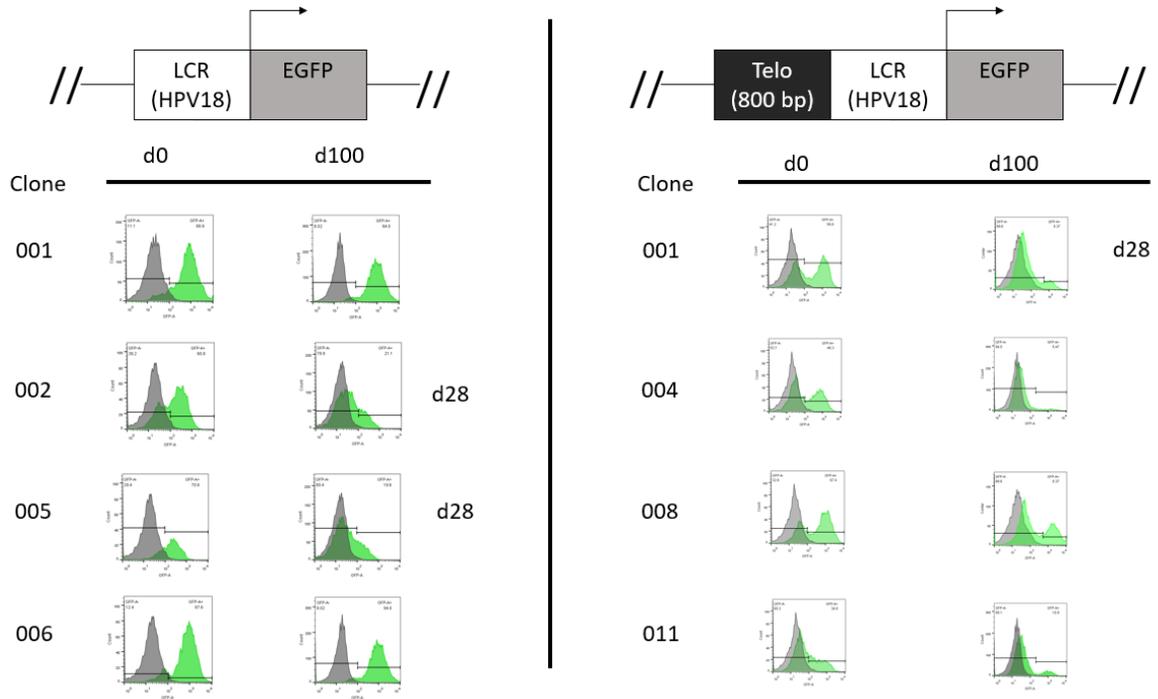


Figura 19. Ensayos de citometría de flujo en clonas obtenidas de las poblaciones transfectadas con los plásmidos pEGFP1-LCR18 y pEGFP1-LCR18-Telo respectivamente. En gris se muestra a la fluorescencia obtenida con células HeLa, sin transfectar y en verde la fluorescencia de las clonas transfectadas. Con base en que los resultados de la fluorescencia obtenida con la línea celular HeLa y la transfectada con pEGFP1 (sin LCR) eran los mismos, se decidió usar HeLa sin transfectar como control para los ensayos de citometría de flujo. En el caso del transgén LCR-EGFP, a los 28 días existen clonas negativas (002 y 005), mientras que a los 100 días postransfección existen clonas positivas (001 y 006). En el caso del transgén Telo-LCR-EGFP, a los 100 días existen clonas negativas (001, 004 y 011) y la clona 008 presenta a los 100 días una población negativa y otra positiva a la actividad de EGFP.

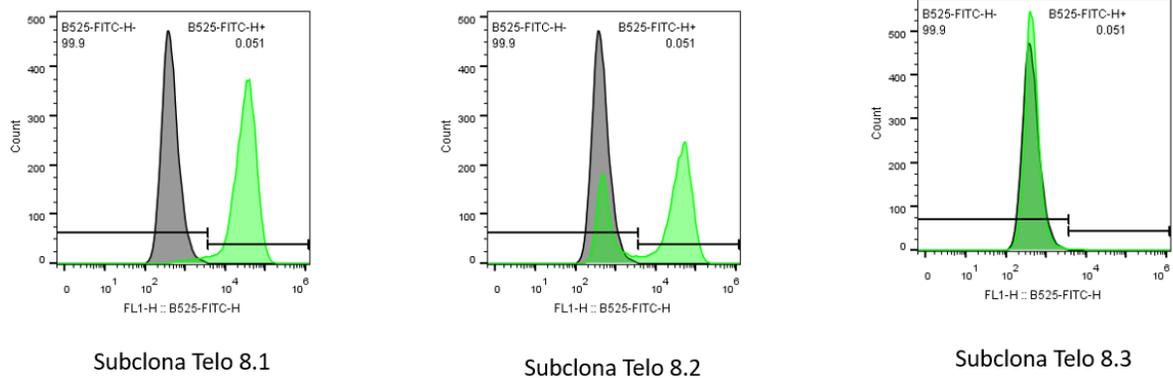


Figura 20. Debido a que la clona 008 mostró poblaciones diferenciales a la actividad de EGFP, se procedió a obtener subclonas de la clona 008 transfectada con el plásmido pEGFP1-LCR18-Telo. Se observaron 3 comportamientos diferentes. Una clona totalmente positiva a la expresión de EGFP (8.1), una clona que recapitula las dos poblaciones positiva y negativa a la expresión de EGFP de la clona 008 (8.2), respectivamente, y una clona totalmente negativa a la expresión del transgén EGFP (8.3). En gris se representa a la fluorescencia obtenida con la línea celular HeLa, sin transfectar y en verde de la fluorescencia de las clonas transfectadas.

Estos resultados indican que la clona 001, transfectada con el plásmido sin el fragmento telomérico, y las clonas transfectadas con el plásmido que contiene al fragmento telomérico (subclonas 8.1 y 8.2) son capaces de mantener la actividad de EGFP durante al menos 100 días con la excepción de la subclona 8.3. Por esta razón fue necesario realizar ensayos de ChIP-qPCR para evaluar la presencia del factor nuclear CTCF en la región LCR del transgén LCR-EGFP, así como la marca epigenética H3K9me3 relacionada con silenciamiento telomérico y con silenciamiento por efecto de posición cromosomal en estas clonas.

8.4. Evaluación de la presencia de CTCF en el transgén estable en clonas obtenidas de la línea celular HeLa.

Con el objetivo de determinar la presencia del factor nuclear CTCF en la LCR de HPV tipo 18 en el transgén de las clonas obtenidas de la línea celular HeLa, se realizaron ensayos de ChIP-qPCR en aquellos casos particulares de las clonas que mantuvieron la actividad del transgén LCR-EGFP durante más de 100 días. Esto sugiere que no existe una protección generalizada de la LCR en el genoma de las células. Para estos ensayos se utilizó la clona HeLa EGFP 001 (que contiene integrado al transgén LCR18-EGFP) y a las clonas HeLa Telo 8.1, 8.2 y 8.3 (que contienen integrado al transgén LCR18-EGFP-Telo). Para estos ensayos se emplearon un par de oligonucleótidos que alinean en el transgén EGFP, así como en la unión de las secuencias de la LCR y el extremo 5' del transgén EGFP para no obtener productos amplificados provenientes de las copias endógenas de HPV tipo 18 de la línea celular HeLa. Como se muestra en las **(Figuras 20-24)** se observó que en las clonas CTCF no se encuentra enriquecido en el transgén y que existe una relación directa entre el silenciamiento del transgén y el enriquecimiento de la modificación postraduccional H3K9me3, relacionada a silenciamiento telomérico o bien puede tratarse de un efecto de posición cromosomal debido al inserto telomérico de la construcción plásmidica en regiones no teloméricas del genoma. Como control positivo para el ChIP del factor nuclear CTCF, se utilizó al gen WRAP53 y como control negativo de la técnica se empleó al exón 27 del gen RB1¹⁰⁴.

Análisis de la ocupación de CTCF y de la marca epigenética H3K9me3 en el transgén LCR-EGFP en la clona HeLa EGFP

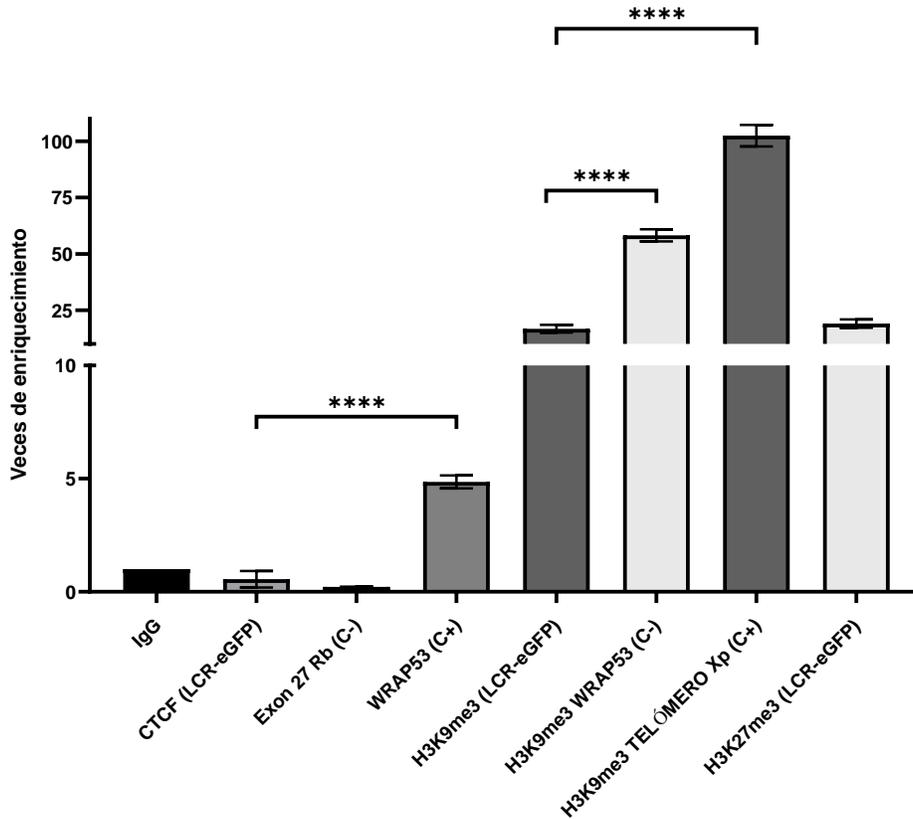


Figura 21. Asociación de CTCF y la marca epigenética H3K9me3 en el transgén LCR-EGFP en la clona HeLa EGFP. Ensayo de ChIP-qPCR para determinar la presencia de CTCF evaluando el transgén LCR18-EGFP en la clona 001 obtenida de la línea celular HeLa. Se presenta el enriquecimiento contra IgG. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). El anticuerpo inespecífico IgG se empleó como control negativo de la técnica de ChIP. El análisis de PCR cuantitativa (qPCR) se realizó utilizando DNA obtenido del ensayo de ChIP que evalúa la presencia del factor nuclear CTCF y de la marca epigenética H3K9me3 en la región comprendida entre la LCR de HPV 18 y el gen EGFP del transgén integrado (LCR-EGFP) en esta clona, así como dos regiones genómicas control para el enriquecimiento de CTCF y H3K9me3. Se observa que en esta clona (HeLa EGFP 001) no se detectó la presencia ni de CTCF ni de la modificación postraduccional H3K9me3 en la LCR del transgén y esto se refleja en que no existe diferencia estadística significativa entre CTCF y su control negativo (WRAP 53) y en un enriquecimiento por debajo del control negativo (WRAP 53) para el caso de H3K9me3. Para CTCF se emplearon como controles negativo y positivo al exón 27 del gen Rb (C-) y al gen WRAP 53 (C+) respectivamente. (Guerra-Calderas, 2018), mientras que para la marca epigenética H3K9me3 se empleó al gen WRAP 53 (C-) como control negativo y como control positivo se empleó la región telomérica del brazo p del cromosoma X (C+).

Análisis de la ocupación de CTCF y de la marca epigenética H3K9me3 en el transgén LCR-EGFP en la clona HeLa Telo 8.1

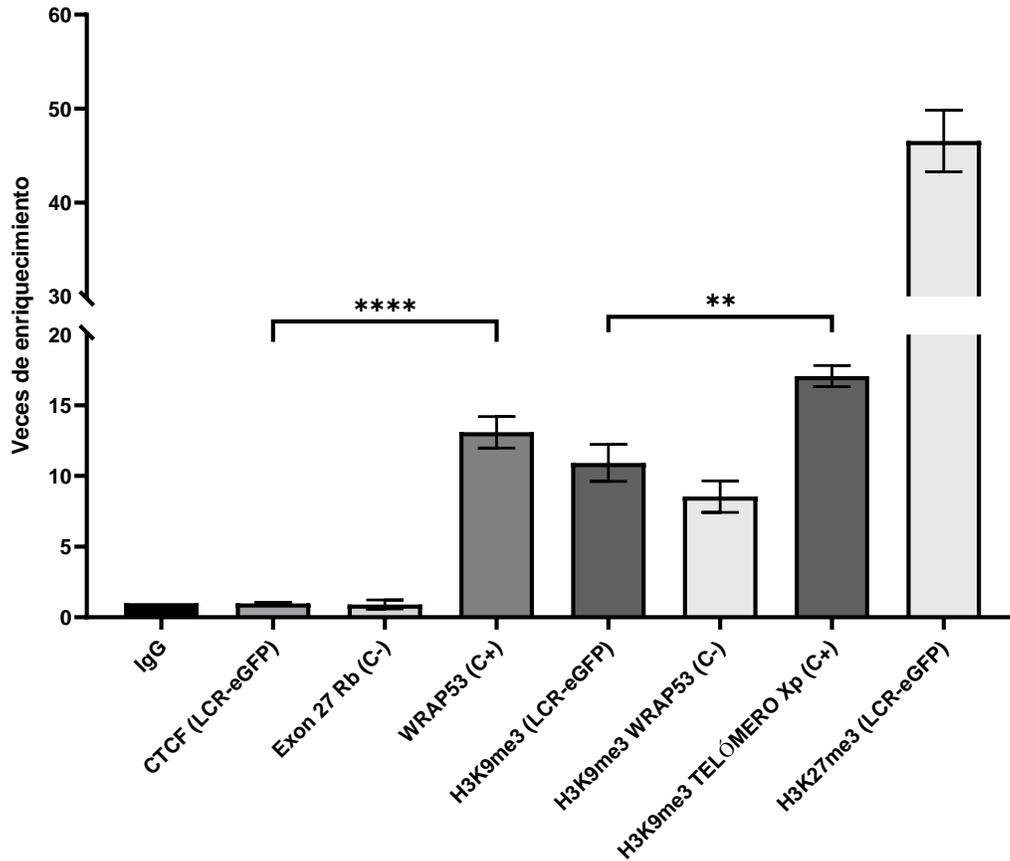


Figura 22. Asociación de CTCF en el transgén LCR-EGFP en la clona HeLa Telo (Clona 8.1). Ensayo de ChIP-qPCR para determinar la presencia de CTCF evaluando el transgén LCR18-EGFP en la clona HeLa Telo 8.1 obtenida de la línea celular HeLa. Se presenta el enriquecimiento contra IgG. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). El anticuerpo inespecífico IgG se empleó como control negativo de la técnica de ChIP. El análisis de PCR cuantitativa (qPCR) se realizó utilizando DNA obtenido del ensayo de ChIP que evalúa la presencia del factor nuclear CTCF y de la marca epigenética H3K9me3 en la región comprendida entre la LCR de HPV 18 y el gen EGFP del transgén integrado (LCR-EGFP) en esta clona, así como dos regiones genómicas control para el enriquecimiento de CTCF y H3K9me3. Se observa que en esta clona (HeLa Telo 8.1) no se detectó la presencia ni de CTCF ni de la modificación postraduccional H3K9me3 en la LCR del transgén y esto se refleja en que no existe diferencia estadística significativa entre CTCF y H3K9me3 con su control negativo (WRAP 53). Para CTCF se emplearon como controles negativo y positivo al exón 27 del gen Rb (C-) y al gen WRAP 53 (C+) respectivamente. (Guerra-Calderas, 2018), mientras que para la marca epigenética H3K9me3 se empleó al gen WRAP 53 (C-) como control negativo y como control positivo se empleó la región telomérica del brazo p del cromosoma X (C+).

Análisis de la ocupación de CTCF y de la marca epigenética H3K9me3 en el transgén LCR-EGFP en la clona HeLa Telo 8.2

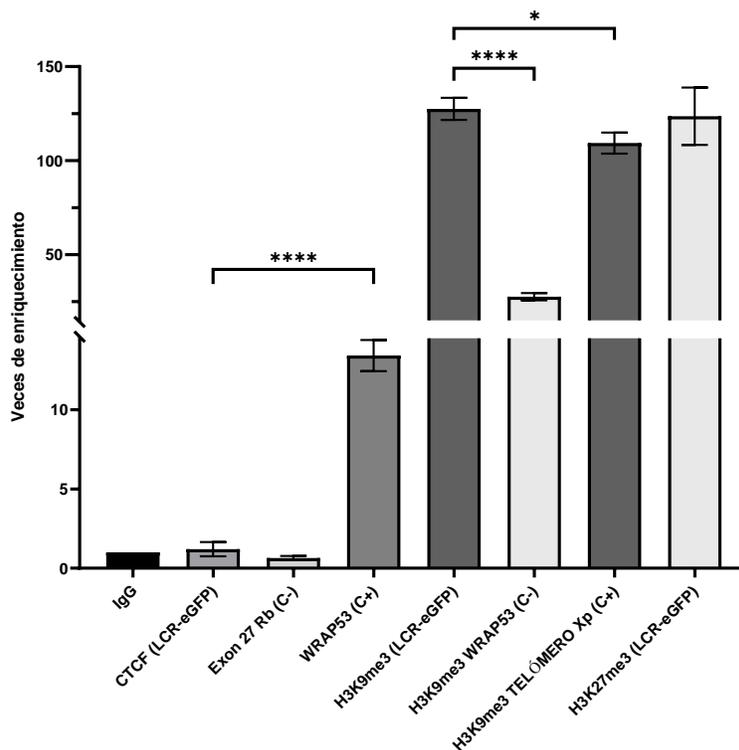


Figura 23. Asociación de CTCF en el transgén LCR-EGFP en la clona HeLa Telo (Clona 8.2). Ensayo de ChIP-qPCR para determinar la presencia de CTCF evaluando el transgén LCR18-EGFP en la clona HeLa Telo 8.2 obtenida de la línea celular HeLa. Se presenta el enriquecimiento contra IgG. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). El anticuerpo inespecífico IgG se empleó como control negativo de la técnica de ChIP. El análisis de PCR cuantitativa (qPCR) se realizó utilizando DNA obtenido del ensayo de ChIP que evalúa la presencia del factor nuclear CTCF y de la marca epigenética H3K9me3 en la región comprendida entre la LCR de HPV 18 y el gen EGFP del transgén integrado (LCR-EGFP) en esta clona, así como dos regiones genómicas control para el enriquecimiento de CTCF y H3K9me3. Se observa que en esta clona (HeLa Telo 8.2) no se detectó la presencia de CTCF en la LCR del transgén y esto se refleja en que no existe diferencia estadística significativa entre CTCF y su control negativo (WRAP 53). Lo que sí se observa es el enriquecimiento de la marca postraduccional H3K9me3 en el transgén LCR-EGFP en esta clona 5 veces por encima del enriquecimiento del control negativo (WRAP 53). Este enriquecimiento es incluso mayor que el obtenido con el control positivo (Telómero Xp). Para CTCF se emplearon como controles negativo y positivo al exón 27 del gen Rb (C-) y al gen WRAP 53 (C+) respectivamente. (Guerra-Calderas, 2018), mientras que para la marca epigenética H3K9me3 se empleó al gen WRAP 53 (C-) como control negativo y como control positivo se empleó la región telomérica del brazo p del cromosoma X (C+).

Análisis de la ocupación de CTCF y de la marca epigenética H3K9me3 en el transgén LCR-EGFP en la clona HeLa Telo 8.3

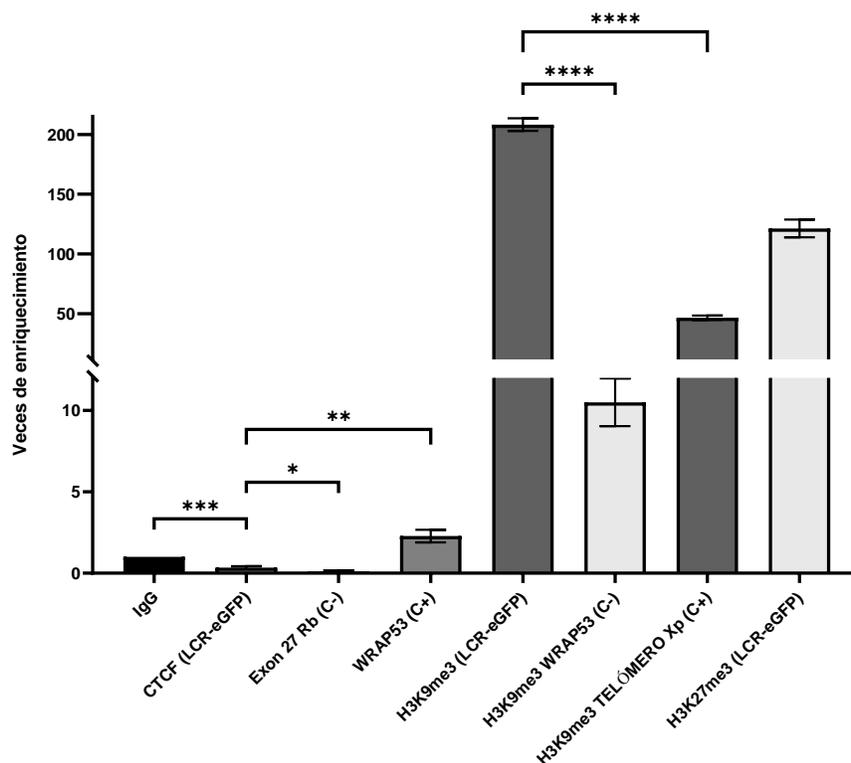


Figura 24. Asociación de CTCF en el transgén LCR-EGFP en la clona HeLa Telo (Clona 8.3). Ensayo de ChIP-qPCR para determinar la presencia de CTCF evaluando el transgén LCR18-EGFP en la clona HeLa Telo 8.3 obtenida de la línea celular HeLa. Se presenta el enriquecimiento contra IgG. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). El anticuerpo inespecífico IgG se empleó como control negativo de la técnica de ChIP. El análisis de PCR cuantitativa (qPCR) se realizó utilizando DNA obtenido del ensayo de ChIP que evalúa la presencia del factor nuclear CTCF y de la marca epigenética H3K9me3 en la región comprendida entre la LCR de HPV 18 y el gen EGFP del transgén integrado (LCR-EGFP) en esta clona, así como dos regiones genómicas control para el enriquecimiento de CTCF y H3K9me3. Se observa que en esta clona (HeLa Telo 8.3) no se detectó la presencia de CTCF en la LCR del transgén y esto se refleja en que no existe diferencia estadística significativa entre CTCF y su control negativo (WRAP 53). Lo que sí se observa es el enriquecimiento de la marca postraduccional H3K9me3 en el transgén LCR-EGFP en esta clona 20 veces por encima del enriquecimiento del control negativo (WRAP 53). Este enriquecimiento es incluso 4 veces mayor que el obtenido con el control positivo (Telómero Xp). Para CTCF se emplearon como controles negativo y positivo al exón 27 del gen Rb (C-) y al gen WRAP 53 (C+) respectivamente. (Guerra-Calderas, 2018), mientras que para la marca epigenética H3K9me3 se empleó al gen WRAP 53 (C-) como control negativo y como control positivo se empleó la región telomérica del brazo p del cromosoma X (C+).

Análisis de la ocupación de H3K9me3 en el transgén LCR-EGFP

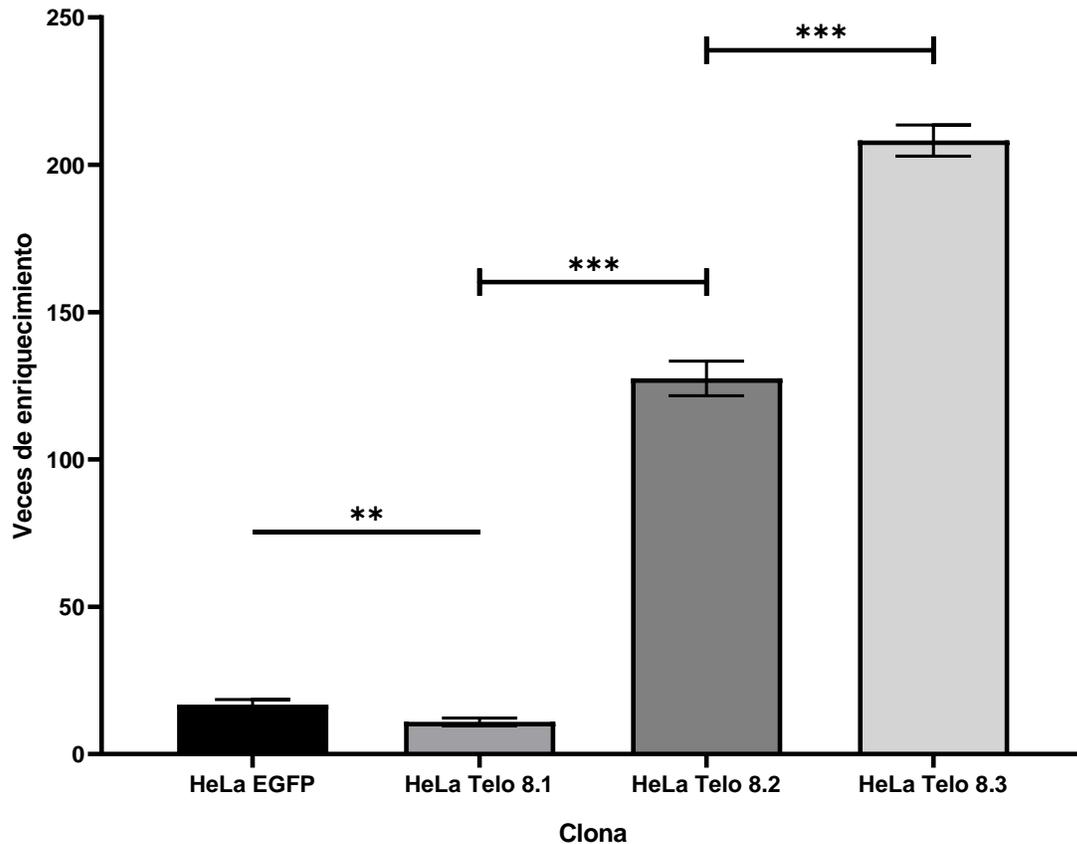


Figura 25. Análisis de la asociación de la marca epigenética H3K9me3 en la LCR del transgén de las clonas HeLa EGFP, HeLa Telo 8.1, HeLa Telo 8.2 y HeLa Telo 8.3. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). Se observa el enriquecimiento de la marca postraduccional H3K9me3 en cada una de las clonas obtenidas. La clona HeLa Telo 8.1 tiene el enriquecimiento más bajo de todas seguida por la clona HeLa EGFP (ambas clonas son 100% positivas a la actividad de EGFP). La clona 8.2 posee un enriquecimiento de aproximadamente 120 por encima del anticuerpo irrelevante IgG y posee poblaciones positivas y negativas a la actividad de EGFP evaluadas mediante citometría de flujo. Finalmente la clona HeLa Telo 8.3 es la más enriquecida en esta marca con un enriquecimiento promedio de 200 por encima del anticuerpo irrelevante IgG y contiene una población totalmente negativa a la expresión de EGFP.

Análisis de la ocupación de H3K27me3 en el transgén LCR-EGFP

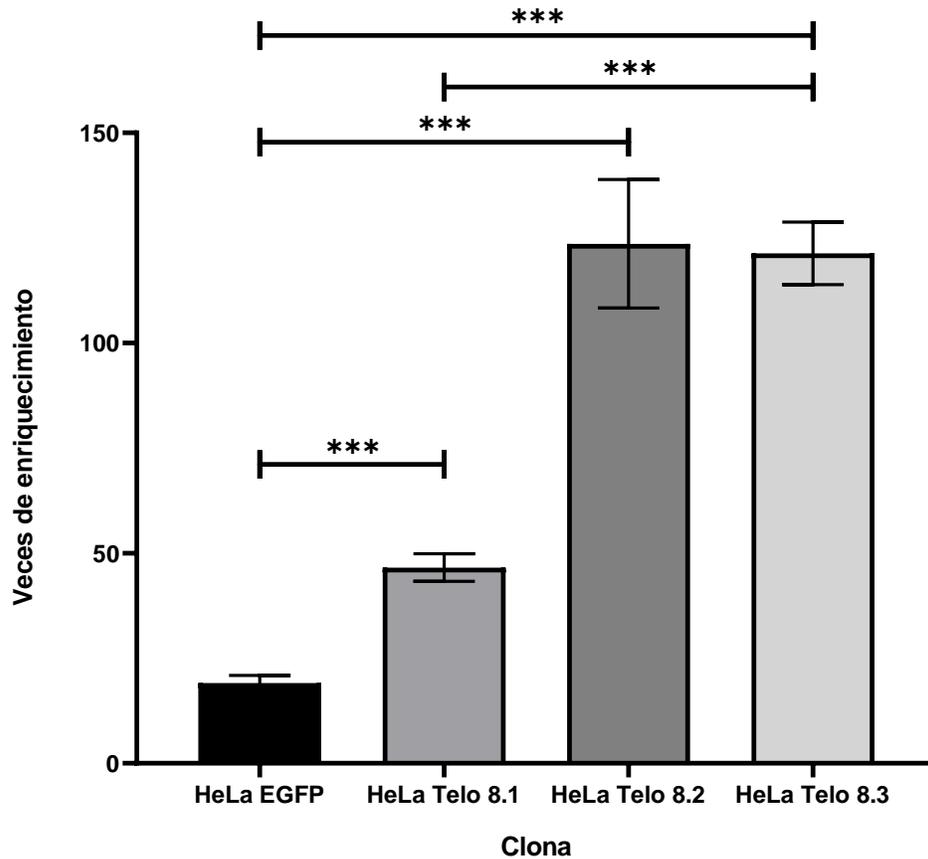


Figura 26. Análisis de la asociación de la marca epigenética H3K27me3 en la LCR del transgén de las clonas HeLa EGFP, HeLa Telo 8.1, HeLa Telo 8.2 y HeLa Telo 8.3. Las diferencias estadísticas significativas se calcularon utilizando una prueba t. Se muestra la media con su respectiva desviación estándar de las mediciones de qPCR por triplicado técnico. Valor de P: 0.1234 (ns), 0.033 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****). Se observa el enriquecimiento de la marca postraducciona H3K27me3 en cada una de las clonas obtenidas. La clona HeLa EGFP tiene el enriquecimiento más bajo de todas seguida por la clona HeLa Telo 8.1 de aproximadamente 50 veces por encima del anticuerpo irrelevante IgG (ambas clonas son 100% positivas a la actividad de EGFP). La clona HeLa Telo 8.2 y 8.3 poseen un enriquecimiento de aproximadamente 120 por encima del anticuerpo irrelevante IgG y mientras que la clona HeLa Telo 8.2 posee poblaciones positivas y negativas a la actividad de EGFP evaluadas mediante citometría de flujo, la clona HeLa Telo 8.3 es. Finalmente la clona HeLa Telo 8.3 contiene una población totalmente negativa a la expresión de EGFP.

Estos resultados indican que CTCF no se une a la región LCR y no participa en la protección del transgén integrado en ninguna de las clonas y que las clonas que tienen poblaciones sin actividad de EGFP existe un enriquecimiento neto de la marca H3K9me3 y un enriquecimiento aparente de la marca H3K27me3 en el transgén, relacionada fuertemente con represión transcripcional, que evita la expresión del gen EGFP. Es muy importante terminar los ensayos de qPCR con los oligonucleótidos controles para la marca H3K27me3. Además, la clona HeLa Telo 8.2 que tiene dos poblaciones una negativa y otra positiva a la actividad de EGFP tiene un menor enriquecimiento de la marca H3K9me3 que la clona HeLa Telo 8.3, que se encuentra completamente apagada pero que mantiene integrado en su genoma al transgén. Curiosamente, los niveles de H3K27me3 son iguales para estas dos clonas (**Figuras 25 y 26**). Las clonas que tienen actividad EGFP no mostraron enriquecimiento de la marca H3K9me3 por encima de su control negativo de qPCR. El mecanismo por el cual permanecen activas debe ser independiente de la participación de CTCF a la LCR en este sistema por lo que deben existir otros factores transcripcionales celulares que se unen a esta región y que la protegen del silenciamiento epigenético y que este silenciamiento está relacionado con la ganancia de las marcas H3K9me3 y H3K27me3 en el transgén.

9. DISCUSIÓN

Este trabajo buscó estudiar la importancia que pudieran tener los sitios de unión para el factor nuclear CTCF dentro de la región larga de control (LCR) de HPV tipo 18. La región LCR, es la región promotora del virus del papiloma humano tipo 18 y es fundamental para el mantenimiento de la actividad transcripcional de este virus, después de su inserción en las distintas regiones de del genoma, incluyendo las regiones de heterocromatina telomérica⁶⁰. Nuestro sistema permitió obtener clonas transfectadas establemente con el transgén LCR con una actividad de EGFP que permanece a lo largo de 100 días. Las limitaciones más importantes de nuestro sistema fueron que muchas de las células transfectadas y seleccionadas establemente resultaron negativas a la actividad de EGFP por lo que en estas células el sistema no permitió la evaluación de la LCR del transgén. Este resultado sugiere que la actividad de la LCR depende del sitio y del número de copias que se integran y del lugar donde se integran y que no existe una protección generalizada, ni por parte de CTCF ni por ningún otro factor transcripcional, de la LCR en la línea celular HeLa, además que de momento no se han hecho los estudios que permitan saber la localización ni el número de copias integradas en las células transfectadas establemente ni tampoco si al integrarse en regiones repetidas el fragmento recombina o se amplifica en células derivadas de cáncer con gran inestabilidad genómica como lo es HeLa¹⁰⁵.

Así como los organismos eucariontes han evolucionado para organizar su genoma con distintas estructuras de su cromatina, los virus del papiloma humano (especialmente los de alto riesgo) también lo han hecho para asegurar su expresión y lograr un control fino de la regulación de su genoma después de infectar a los queratinocitos de las capas basales de epitelios estratificados y para poder sincronizarse con el programa de diferenciación de la célula hospedera⁹⁵. En efecto, el genoma de HPV tipo 18 posee nucleosomas bien posicionados que permiten la expresión controlada de los oncogenes E6 y E7, así como una marcada susceptibilidad al corte por DNasa I en la región larga de control (LCR)¹⁰⁶. En estado episomal, se han demostrado varios sitios de unión para el factor nuclear CTCF en el genoma de HPV tipo 18, siendo el más importante el que se encuentra dentro del ORF del gen E2¹⁰⁷. Así mismo, se ha demostrado que al eliminar este sitio, episomalmente se pierden las interacciones de CTCF con el genoma episomal de HPV tipo 18¹⁰⁸.

Este trabajo permitió evaluar y demostrar en un modelo de cultivo celular *in vitro* la unión del factor nuclear CTCF a la LCR endógena de HPV tipo 18 en células HeLa. Lo cual sugería que el factor nuclear CTCF pudiera tener funciones importantes en este tipo de elementos reguladores virales. Sin embargo, es importante mencionar que la unión de CTCF en la LCR de HPV 18 endógeno, es decir, una vez integrado parece no ser un caso general para las más de 50 copias de esta secuencia integradas en el genoma de la línea celular HeLa. Ahora se conoce que CTCF se une al ORF del gen E2 de HPV en su forma episomal, es decir no integrada, y que la unión de CTCF en el ORF del gen E2 viral es fundamental para que CTCF pueda unirse a otras regiones del genoma del virus, particularmente en la LCR¹⁰⁸. Mutar el sitio de unión a CTCF en el gen E2 no tiene efectos en la replicación del HPV, pero sí en la producción de transcritos virales¹⁰⁹. También se sabe que, al integrarse, el gen E2 queda interrumpido de tal forma que el HPV no se integra completo, pero existen casos donde una parte de la secuencia del gen E2 se integra, particularmente la parte que contiene el sitio de unión a CTCF, y a pesar de la integración de esta secuencia con el sitio de unión, no es posible detectar a CTCF en este sitio¹⁰⁹. Debido a que existen más de 50 copias de esta secuencia en las células HeLa, este resultado también pudiera ser explicado a través de eventos de integración cercanos a regiones genómicas a las que sí pueda unirse CTCF¹¹⁰. Otro fenómeno que discutir es el hecho de que los controles utilizados no se encuentran en la misma dosis que el número de copias que la LCR, como es el caso de los controles WRAP53 y RB1 que se encuentran 8 copias en el genoma versus las 50 copias de la LCR de HPV endógeno, y que a pesar de tener un enriquecimiento de CTCF en la LCR endógena, sería válido pensar que muy seguramente CTCF no se encuentra presente en la LCR de la mayoría o de todas las copias de HPV. Otra serie de experimentos diseñados para saber si realmente CTCF se une en esta región genómica serían experimentos de ChIP-seq, resonancia de plasmones superficiales y ensayos de captura de conformación de cromosomas (3C).

Dado que una de las múltiples funciones de CTCF es precisamente la de evitar la propagación de la heterocromatina en transgenes integrados en los telómeros de los cromosomas⁵⁸, este trabajo inicialmente buscó identificar esta función protectora de CTCF como la responsable del mantenimiento de la actividad promotora de la LCR¹¹¹. Esta propiedad de CTCF era de gran interés para este proyecto debido a que existen reportes que localizan copias del genoma de HPV en los telómeros de los cromosomas de carcinomas cervicales¹¹² mediante RNA-seq lo que indicaba que estas copias son copias activas de HPV que se localizan en estas regiones y que siguen expresando a los

oncogenes E6 y E7⁶⁰. Esto sugiere la existencia de mecanismos que protegen a la LCR del silenciamiento inducido por las secuencias repetidas de estas regiones teloméricas. Dado que estas regiones se encuentran fuertemente reprimidas, este hallazgo toma mucha importancia porque, una vez elucidado, este mecanismo podría emplearse para inducir el silenciamiento de copias activas de HPV integradas en este tipo de regiones y pudiera emplearse como tratamiento de pacientes con carcinoma cervical.

De nuestra evaluación *in silico* de la región larga de control (LCR) del HPV tipo 18, se pudieron identificar 2 posibles sitios de unión para CTCF con los scores más altos del programa dentro de la región 5' de la LCR, así como en la región proximal. Esto quiere decir que el HPV, así como otros muchos virus de DNA, ha evolucionado para utilizar a CTCF con la finalidad de cumplir eficientemente su ciclo de vida⁸⁴ y que los HPVs, en particular los de alto riesgo, han evolucionado para reclutar a CTCF a su genoma para regular la expresión de los oncogenes virales E6 y E7 en el contexto de una infección productiva¹⁰⁷.

La integración no es parte del ciclo de vida de los HPVs y la integración del genoma del HPV tipo 18 en el genoma de las células propone un mecanismo alternativo relevante para la carcinogenicidad producida por la integración del HPV que estaría relacionada con la integración de un enhancer (ensayos de ChIP-seq en la línea celular HeLa reportan la presencia de marcas epigenéticas relacionadas con elementos tipo enhancer¹¹³, H3K4me1 y H3K27ac, en la LCR de HPV tipo 18 integradas en el genoma de estas células¹¹³) que tendría la capacidad de contactar con otros sitios del genoma a través de su sitio CTCF y esto podría resultar en la activación de oncogenes¹¹⁴. En efecto, se ha demostrado que en la línea celular HeLa una de las integraciones de HPV tipo 18 se localiza a 500 kb del protooncogén MYC y por medio de ensayos de 3C se observó que estas regiones interactúan y gracias a esta interacción, MYC se sobreexpresa gracias al HPV¹¹⁴. Actualmente se piensa que este fue el evento inicial de la tumorigenesis en estas células¹¹⁴. Interesantemente, al eliminar el fragmento de HPV integrado cerca del gen MYC por el sistema CRISPR / Cas9, la expresión de MYC se reduce drásticamente en células HeLa¹¹⁵.

Para corroborar que nuestras observaciones *in silico* de la LCR de HPV 18 pudieran reproducirse en ensayos *in vitro* y utilizando los ensayos de ChIP-qPCR en la línea celular HeLa, observamos que el factor nuclear CTCF se encuentra unido en la LCR de las copias integradas del virus HPV tipo 18, presente en esta línea celular. Esto parecía indicar que la asociación del CTCF al LCR se lleva a cabo de manera natural. Ciertamente esta asociación ocurre de forma natural en el contexto de la infección productiva de HPV donde el virus nunca se integra en el genoma de las células y es capaz de avanzar a través de los epitelios estratificados desde la capa basal hasta los estratos más queratinizados y produce viriones que pueden infectar a otras células¹⁰⁸. Ciertamente, los resultados del ChIP sugieren que existe esta unión en estas células, sin embargo, como ya se mencionó anteriormente, este resultado debe interpretarse cuidadosamente debido a la diferencia de la dosis de la LCR con respecto a la de los controles y esto podría generar un falso positivo. De existir esta unión, esto tendría repercusiones biológicas y patológicas importantes en la infección con HPV debido a que la modulación del reclutamiento de CTCF en la LCR podría tener resultados muy distintos dependiendo del momento de la infección con HPV¹⁰⁸. Esto es porque en las primeras etapas de la infección con HPV, en queratinocitos basales y en la que no existe integración del genoma viral, CTCF se encarga de reprimir la expresión de los oncogenes virales E6 y E7 debido a que estas células se siguen dividiendo y no se necesita de estas oncoproteínas para favorecer este proceso¹⁰⁸. Esto significa que interrumpir la unión de CTCF en la forma episomal de HPV en las primeras etapas de la infección podría incluso favorecer la aparición del cáncer. Esto quiere decir que en las primeras etapas de la infección debería más bien buscarse la manera de que CTCF permanezca unido a la LCR para impedir la expresión de los oncogenes en los estratos superiores de los epitelios. Una vez integrado, se pierde parte del genoma viral y CTCF ya no se puede unir al genoma viral integrado. Efectivamente, datos de ChIP-seq en células HeLa indican que esta unión se pierde definitivamente incluso cuando la secuencia del HPV que se integró contiene la parte del ORF del gen E2 que contiene al sitio de unión a CTCF⁹⁵. La integración del HPV finalmente favorece la expresión de los oncogenes E6 y E7 y se considera como un evento necesario para la aparición del cáncer¹¹⁶. Estrategias distintas que podrían proponerse para saber si realmente esta unión ocurre de manera natural sería realizar ensayos de ChIP-seq, DamID, CLIP o CLIP-seq.

Para comprobar si el reclutamiento de CTCF por la LCR de HPV18 era capaz de afectar la expresión génica de un transgén, se diseñaron diferentes construcciones plasmídicas que permiten insertar al gen reportero EGFP bajo el control de la LCR de HPV tipo 18 en la línea celular HeLa. Se eligió esta línea celular debido a que se sabe que la actividad de la LCR del HPV depende del tipo celular que infecta⁶⁶. La secuencia del fragmento telomérico consistió de 800 pb del repetido TTAGGG, y esta estrategia ha sido empleada anteriormente para dirigir la integración de transgenes hacia los telómeros de los cromosomas^{103,117}. Los resultados de las transfecciones fueron muy interesantes, ya que las células transfectadas con el plásmido sin el fragmento telomérico, mostraron una subpoblación del 44% positiva para la expresión de la EGFP, mientras que las células transfectadas con el plásmido que contiene al repetido telomérico sólo mostraron una población positiva a EGFP del 16.9 %. Esto puede deberse a que el transgén que contiene el repetido telomérico está siendo dirigido hacia los telómeros de los cromosomas otra alternativa es que la secuencia por sí sola al ser un repetido telomérico tiene en efecto sobre el transgén y este fenómeno bien pudiera tratarse de un silenciamiento mediado por efecto de posición cromosomal inducido por el fragmento telomérico¹¹⁸. Como se mencionó en los resultados, se obtuvieron poblaciones homogéneas para poder medir la actividad del LCR-EGFP a lo largo de 100 días, mediante el método de clonación por dilución seriada. Se realizaron ensayos de citometría de flujo en las clonas obtenidas a lo largo de 100 días, con el fin de evaluar a la LCR del transgén a través del seguimiento de la actividad de EGFP. Los resultados fueron variados: mientras que las clonas 001 y 006, que no contenían el fragmento telomérico, mantuvieron la expresión de la EGFP, a lo largo de 100 días, las clonas 002 y 005, que tampoco contenían clonado al fragmento telomérico, comenzaron a perder la actividad de expresión de la EGFP dentro de los primeros 28 días, mostrando una expresión menor al final de este periodo. Esto dejó claro algunas de las limitaciones más importantes de nuestro modelo siendo que el 56% de las células transfectadas con el plásmido sin fragmento telomérico se encontraban apagadas después de 48h y las que tenían actividad EGFP comenzaban a perderla en los primeros 28 días postransfección. Además, no era posible evaluar a la LCR a través de la actividad de EGFP en aquellas clonas que tuvieron silenciado el transgén. Esto es importante porque incluso sin fragmento telomérico que favorezca la integración de la LCR hacia estas regiones, las células parecen tener la capacidad de silenciar a la mayoría de los transgenes integrados por lo que vuelve muy interesante conocer donde se integraron las copias activas del transgén y que mecanismo utilizaron para evitar el silenciamiento inducido por la maquinaria celular. Por

otro lado, con respecto a las clonas que sí contenían al fragmento telomérico, las clonas 001 y 004 perdieron la expresión de la EGFP al día 100, mientras que las clonas 008 y 011 mantuvieron parcialmente la expresión de la EGFP y al final de este tiempo se pudieron observar dos poblaciones bien definidas, una positiva y otra negativa a la expresión de EGFP. Con el fin de entender este comportamiento se subclonó a clona 008 para la obtención de clonas por el método de dilución seriada. La subclona 8.1 resultó ser completamente positiva a la expresión de EGFP; mientras que la subclona 8.2 volvió a comportarse como la primera clona 008. Es decir, se generó una subpoblación positiva y otra negativa a la expresión de EGFP. Por otro lado, encontramos a la subclona 8.3, la cual resultó ser completamente negativa a la expresión del transgén. Estos resultados nos podrían indicar que en la clona 008 se estuvieron generando modificaciones epigenéticas y genéticas que dieron como resultado estas distintas subpoblaciones celulares. Lo cual, podría estar hablando de la dinámica que puede generarse en las distintas subpoblaciones celulares *in vitro*. Desde el punto de vista genético, pueden haber varias copias del transgén en todo el genoma de las clonas, lo que explicaría porque algunas conservan la actividad de EGFP, también se sabe que la LCR tiende a integrarse en tándem y esto puede favorecer la expresión de EGFP o bien, pudieran estarse dando eventos de recombinación del fragmento telomérico del transgén con los telómeros de los cromosomas que ocasione que se transloquen estas secuencias y se pierda la actividad de EGFP y una vez retirada la selección con G418 la presión selectiva que mantiene encendido el gen de resistencia a G418 desaparece y el transgén se silencia. Desde el punto de vista epigenético, se sabe que metilación del DNA en la LCR de HPV es importante para su actividad transcripcional por lo que en algunas subpoblaciones de la clona 008 pudo haberse metilado el transgén y perder la actividad de EGFP¹¹⁹. Al obtener las subclonas, este fenómeno recapitula y observamos clonas sin actividad de EGFP. Otro evento epigenético importante sería el de efecto de posición cromosomal, donde los transgenes que se hayan integrado en regiones de heterocromatina quedaron silenciados por este efecto. Se ha observado también que la LCR puede contactar con otras regiones del genoma para favorecer la expresión de oncogenes, por lo que hipotéticamente solo aquellos transgenes que se encuentran contactando otras regiones del genoma son las que pudieran conservar su actividad de EGFP. Para poder comprobar estas hipótesis es fundamental conocer la localización de la o las copias del transgén en estas clonas mediante secuenciación completa de su genoma. También sería muy importante investigar el estado epigenético del genoma de HPV en células HeLa mediante ensayos de ChIP-seq disponibles para saber que marcas

epigenéticas se encuentran enriquecidas en los genomas integrados de HPV en estas células.

En la clona EGFP 001 y en las subclonas Telo 8.1, 8.2 y 8.3 no se observó la presencia del factor nuclear CTCF en el transgén LCR-EGFP, lo cual es consistente con el hecho de que CTCF solamente se une a la LCR en el contexto episomal del HPV y que una vez integrado CTCF ya no puede unirse a esta región y esto forma parte del mecanismo por el cual se desregula la expresión génica del HPV y favorece la aparición del cáncer¹⁰⁹. Lo que sí se observó fue la presencia de las marcas epigenéticas H3K9me3 y H3K27me3 en aquellas clonas en las que existía una población negativa a la expresión del transgén EGFP y estas marcas están asociadas fuertemente al silenciamiento epigenético de heterocromatina constitutiva y facultativa respectivamente¹²⁰. Como se mencionó anteriormente, se ha demostrado que en estado episomal, la unión del factor nuclear CTCF en la secuencia del gen viral E2 es fundamental para que CTCF pueda unirse a la LCR de HPV tipo 18 y para poder formar un asa cromatínica junto con el factor YY1 para evitar la expresión de los oncogenes E6 y E7 en las primeras etapas de la infección con este virus¹⁰⁸. Cuando se forma esta estructura, el genoma de HPV se ve enriquecido en la marca H3K27me3 y esta parece ser la forma en la cual se regula la transcripción viral en estado episomal durante la infección productiva con HPV¹⁰⁸.

La importancia y novedad de estos resultados radica en que, una vez integrado en el genoma de la célula, el mecanismo por el cual se puede estar regulando la transcripción viral es a través del enriquecimiento de la marca H3K9me3 además del enriquecimiento de la marca H3K27me3 que regula la transcripción en el estado episomal del virus¹⁰⁸.

¿A quién pudiéramos proponer como el factor protector de la actividad de la LCR? A YY1. YY1 es una proteína arquitectónica que se une a enhancers, promotores y puede formar dímeros para promover interacciones entre estos elementos¹⁰⁸. En efecto, se sabe que la LCR de HPV18 posee múltiples sitios de unión caracterizados para el factor YY1 que junto con CTCF se encarga de regular la transcripción de los oncogenes virales en estado episomal. YY1 disminuye a medida que los queratinocitos basales se diferencian y migran hacia las capas superiores de los tejidos estratificados, pero se ha encontrado a YY1 sobre expresado en una gran cantidad de cánceres, incluyendo al cáncer cervicouterino¹²¹. YY1 es un factor que también tiene la capacidad de reclutar a la maquinaria (PRC2) que deposita la marca epigenética H3K27me3 y resulta interesante pensar que este factor pudiera ser el responsable de este comportamiento dual en la LCR de HPV tipo 18 donde los YY1 que

reclutan a PRC2 silencian a los transgenes mientras que los YY1 que forman asas de cromatina los mantienen encendidos¹²². Por esta razón sería importante realizar ensayos de ChIP-qPCR para el factor YY1 para comprobar esta hipótesis.

Finalmente, debido a que HeLa tiene aproximadamente 50 copias de HPV en su genoma esto representa una limitante del sistema que compite por los mismos factores transcripcionales que el transgén LCR-EGFP. También es probable que un elevado número de copias ocasione el silenciamiento del transgén como sucede en la línea celular CasKi, que posee más de 600 copias de HPV 16 en su genoma, pero en esta línea celular solo una copia es activa¹¹⁹. Un sistema más adecuado para el estudio de la actividad promotora de la LCR sería mediante la transfección de queratinocitos con un plásmido que contenga a la LCR controlando la expresión de los oncogenes E6 y E7. Se podría emplear el sistema CRISPR-Cas para dirigir al transgén LCR/E6/E7 a regiones específicas del genoma y así evitar múltiples integraciones del transgén. De esta manera, únicamente se seleccionarían los queratinocitos que expresen estos oncogenes y se inmortalicen. De esta forma las clonas que se obtengan serían todas positivas a la expresión de los oncogenes y mediante ensayos de secuenciación del genoma completo podría saberse la localización de estos transgenes en el genoma de las células. Para conocer que copias de HPV son activas se podrían realizar ensayos de RNA FISH.

10. CONCLUSIONES

1. CTCF se une en un modelo *in vitro* a la región larga de control (LCR) endógena del HPV tipo 18 en la línea celular HeLa.
2. En nuestro sistema diseñado para poder estudiar a la LCR de HPV18 mediante el seguimiento de la actividad de la proteína verde fluorescente (EGFP) no fue posible detectar la presencia del factor nuclear CTCF en la LCR de HPV tipo 18 y por lo tanto CTCF no participó en prevenir el silenciamiento epigenético de la LCR de HPV tipo 18 mediado por la maquinaria de represión relacionada con silenciamiento telomérico y con silenciamiento por efecto de posición cromosomal.
3. Se obtuvieron clones de la línea celular HeLa que mantuvieron la expresión del transgén EGFP a pesar de tener un fragmento telomérico vecino lo que indica la existencia de mecanismos independientes de CTCF que protegen a la LCR del silenciamiento epigenético inducido por esta secuencia.
 - a. La integración estable del transgén LCR18-EGFP y LCR18-EGFP-Telo dentro del genoma celular conserva su actividad transcripcional durante un periodo de al menos 100 días lo que deja abierta la posibilidad de que existan otros mecanismos que estén participando en este fenómeno.

11. PERSPECTIVAS

1. Realizar los ensayos de ChIP-qPCR para el factor nuclear CTCF y para las marcas postraduccionales de histonas H3K9me3 y H3K27me3 en las clonas HeLa EGFP 001, HeLa Telo 8.1, 8.2 y 8.3 para así obtener replicas biológicas de estas clonas.
2. Realizar los ensayos de qPCR con oligonucleótidos control para la marca epigenética H3K27me3.
3. Con la finalidad de conocer la ubicación precisa de la integración del transgén LCR18-EGFP, comprobar mediante la secuenciación del genoma completo de cada clona y ensayos de DNA FISH, la localización del transgén en los cromosomas de las células transfectadas.
4. Realizar ensayos de abatimiento de la marca epigenética H3K9me3 y analizar la expresión del gen reportero EGFP a través del tiempo en las clonas transfectadas establemente (clona Telo 8.2 y Telo 8.3).
5. Comprobar mediante ensayos de ChIP el abatimiento de la marca epigenética H3K9me3 mediante inhibidores específicos en las clonas transfectadas establemente y descritas en el punto 2.

12. METODOLOGÍA

Soluciones:

Buffer de entrecruzamiento	Buffer de lisis
Formaldehído al 11 %. NaCl 100 mM. EGTA 0.5 mM. HEPES 50 mM con pH 8.0	SDS al 1 %. EDTA 10 mM con pH 8.0 Tris-HCl 50 mM con pH 8.0 Inhibidores de proteasas.
Glicina 2.5 M	PBS 1x
Pesar 9.38 g de glicina Aforar a 50 mL con agua bidestilada.	8 g NaCl. 0.2 g KCl. 1.44 g Na ₂ HPO ₄ . 0.24 g KH ₂ PPO ₄ . pH=7.4. Aforar a 1 L.

Anticuerpos:

Anticuerpo	Número de Catálogo	Compañía
α -CTCF	07-449	Millipore
α -H3K9me3	C15200146	Diagenode
α -H3K27me3	Ab6002	Abcam

Obtención y evaluación de la cromatina: Las clonas obtenidas de la línea celular HeLa fueron cultivadas en cajas de 150 mm a una confluencia del 80%. Para realizar el entrecruzamiento, a 18 mL de medio DMEM se le agregaron 2mL de buffer de entrecruzamiento para obtener una concentración final de formaldehído al 1%. Las células fueron agitadas durante 10 min con este buffer. Para detener el entrecruzamiento, se agregó 1 mL de glicina 2.5 M para obtener una concentración final de glicina de 125 mM durante 5 min. Las células fueron lavadas 3 veces con PBS frío (4°C). Una vez retirado el PBS, se adicionó 1 mL de buffer de lisis con inhibidores de proteasas a todas las cajas y se incubaron a 4°C durante 10 minutos. Después de este tiempo, las células fueron cosechadas y la cromatina fue transferida a un tubo falcón de 15 mL. Para realizar la sonicación de la cromatina, se colocó el tubo que contiene la cromatina en un vaso con hielo para que se mantenga frío, evitando generar mucha espuma. El programa de sonicación fue de 30 segundos, pulso 2.0 - 0.5, amplitud de 35% en un sonicador Ultrasonic Processor (marca GENEQ, modelo GEX500, SOVC505-00). Para determinar el grado de fragmentación de la cromatina, el tamaño adecuado para las Inmunoprecipitaciones es de 500-700 pb, se tomó una alícuota de 10 uL y se agregaron 5 µL de RNasa A a 37°C durante una hora. Posteriormente se agregaron 5 µL de proteinasa K se incubó a una temperatura de 55°C, por lo menos 2 horas y después a 65°C durante por lo menos 6 h. Subsecuentemente se realizó una extracción con fenol y cloroformo y se precipitó el DNA en una solución de isopropanol y 0.66 volúmenes de acetato de amonio al 12.5 M y se incubó a -20°C durante toda la noche Finalmente, se realizó una electroforesis en un gel de agarosa al 1 % para evaluar el tamaño de la cromatina.

Cuantificación de la cromatina: Para el uso del One Day ChIP Kit se requiere usar la misma cantidad de cromatina para cada IP, para esta cuantificación se procede a determinar la concentración de las cromatinas mediante el método de Lowry.

- a. Preparación de los reactivos: Agregar 20 µL de la solución S por cada mL de la solución A.
- b. Curva estándar de albúmina: Partir de un stock de 1 mg/mL, realizar la curva estándar con las siguientes cantidades de albúmina, 0 mg, 1 mg, 5 mg, 10 mg, 15 mg, 20 mg, con buffer de lisis.
- c. Cargar la placa de ELISA: Con el fin de anular el ruido causado por los detergentes del buffer de lisis, se utiliza la solución S, se agrega 20µL de solución S por cada 1 mL de reactivo A. Por cada muestra agregar 160 µL del reactivo B de Lowry, 20µL del reactivo

A, de 2 μ L a 5 μ L de la muestra, ajustar el volumen final con agua a 200 μ L. Para los volúmenes de la curva estándar se pueden adicionar de 1 μ L a 15 μ L.

ChIP (Inmunoprecipitación de la Cromatina): La técnica de la inmunoprecipitación de la cromatina consiste básicamente en 3 pasos importantes: Inmunoselección, Inmunoprecipitación, Aislamiento y Purificación.

Inmunoselección:

Una vez obtenida la cromatina del tamaño adecuado para las inmunoprecipitaciones correspondientes, se procedió a realizar la inmunoselección. Esta inmunoselección consiste en preparar el Buffer ChIP 1X (a partir de Buffer ChIP 5X One Day ChIP-Diagenode) y agregar los inhibidores de proteasas (200X). Para 20 inmunoprecipitaciones (IPs) se añaden 25 μ L a 5 mL de ChIP Buffer 1X. Ajustar el volumen de manera tal que, 60 μ L de cromatina (correspondiente a 2,000,000 de células) sean diluidos en 220 μ L de Buffer ChIP 1X (con inhibidores de proteasas). El volumen de cromatina no debe ser mayor al 28% de la mezcla total. Una vez diluida, la mezcla se deposita en tubos de 1.5 mL. A cada tubo se le agrega el anticuerpo correspondiente y se incuban toda la noche a 4C en movimiento. Para el INPUT, se toma el mismo volumen de cromatina empleado para las inmunoprecipitaciones y se procesa más adelante.

Inmunoprecipitación:

Una vez terminado el tiempo de incubación de la inmunoselección, los tubos que contienen a las cromatinas con sus respectivos anticuerpos son centrifugadas durante 10 min a una velocidad de 14,000 x g a 4C. Este paso elimina agregados inespecíficos que se depositan en el fondo del tubo. Con cuidado se transfiere el sobrenadante que contiene a los complejos cromatina-anticuerpo y estos son depositados en un tubo nuevo que contiene perlas magnéticas previamente lavadas para poder así aislar estos complejos. Este proceso se realiza incubando los tubos que contienen a los complejos perlas-anticuerpo-cromatina durante 1h a 4C en movimiento. Brevemente, las perlas magnéticas son sometidas a 3 lavados por inmersión en 1 mL de Buffer ChIP 1X con inhibidores de proteasas, aisladas mediante un rack magnético, resuspendidas en Buffer ChIP 1X con inhibidores de proteasas y finalmente depositadas en tubos de 1.5 mL por cada inmunoprecipitación realizada.

Aislamiento:

En este paso se procesan las inmunoprecipitaciones, así como el INPUT.

INPUT: Se agrega 1 mL de EtOH (100%) a la alícuota tomada anteriormente y se incuba durante 10 minutos en frío. Después se centrifuga durante 10 min a una velocidad de 10,000 x g. Se retira el sobrenadante y se realiza un lavado con EtOH (75%). Se realiza una segunda centrifugación de 10 min a una velocidad de 10,000 x g, se retira el sobrenadante y se deja secar el pellet a temperatura ambiente.

Una vez terminada la incubación de las inmunoprecipitaciones con las perlas magnéticas, se utiliza un rack magnético para recuperar las perlas magnéticas y se descarta el sobrenadante sin perturbar el pellet unido al rack magnético. Se realizan 3 lavados con Buffer CHIP 1X y en cada lavado se emplea el rack magnético para recuperar el pellet de perlas magnéticas.

Purificación:

Se agregan 100 uL de la solución purificadora (Slurry) a los tubos que contienen las inmunoprecipitaciones así como al INPUT (a este último previamente se le agregaron 100 uL de agua para PCR y se resuspendió el pellet). Se incuban estos tubos durante 1 min y después se colocan en agua hirviendo durante 10 min. Una vez transcurrido este tiempo, se agregan a todos los tubos 1 uL de RNAsa A (37C, 1h), 1 uL de proteinasa K (65C, 1h) y se incuban a 65C (2h). Después se realiza una segunda incubación en agua hirviendo durante 10 min y una vez completada esta incubación se centrifugan los tubos durante 1 min a 14,000 x g. Se transfieren cuidadosamente, sin perturbar el pellet, 70 uL del sobrenadante a tubos nuevos y se agregan 130 uL de agua grado PCR al pellet para resuspenderlo y volver a centrifugarlo durante 1 min a 14,000 x g. De nuevo se transfieren 130 uL del sobrenadante a los tubos donde fueron depositados los 70 uL anteriores. El volumen final para cada una de las muestras es de 200 uL y estas pueden ser preservadas a -20C. A partir de este momento las muestras pueden ser evaluadas y analizadas mediante la técnica de PCR en tiempo real.

PCR tiempo real (qPCR) utilizando el reactivo SYBR Green

Para esta técnica se emplearon los reactivos Maxima SYBR-Green/ ROX qPCR MasterMix (2X) (Thermo Scientific, #K0222)

Primers:

Oligonucleótido	Secuencia 5'-->3'
LCR-EGFP-1-FWD	GGGACCGAAAACGGTGTAT
LCR-EGFP-1-3-RVS	TGGTGCAGATGAACTTCAGG
EGFP-N	CGTCGCCGTCCAGCTCGACCAG
ORI FWD1	TAGCGCTACCGGACTCAGAT
LCR18-BamHI-FWD	CGGGATCCCGTATGTGTGTGTGTATATATA
LCR18-BamHI-RVS	CGGGATCCCGAGTATTGTGGTGTGTTTC
LCR18 qPCR FWD	TGTTTGTGGTATGGGTGTTGC
LCR18-qPCR-RVS	GTGCCAGCGTACTGTATTGTG

1. Descongelar los reactivos y muestras en hielo.
2. Preparar Master Mix.

Componente	Volumen (uL)/ Reacción
Maxima SYBR GREEN/ ROX qPCR Master Mix (2X)	10.0
Primer FWD	0.5
Primer RVS	0.5
DNA	5.0
NFW (Nuclease Free Water)	4.0
Volumen Final	20.0 uL

3. Colocar los tubos en el equipo termociclador para tiempo real (RotorGene Q, Qiagen) y programar el equipo.

		45 ciclos	
	Desnaturalización	Desnaturalización	Alineamiento (Tm)
Temperatura (°C)	95	95	60
Tiempo	5 min	30 s	30 s (Lectura)

4. Analizar los datos empleando el método $\Delta\Delta C_t$.

Transfecciones: Los plásmidos LCR18-EGFP y LCR18-TELO-EGFP fueron transfectados en la línea celular HeLa de acuerdo con el protocolo sugerido por el fabricante (Xfect, Takara #631318).

Plásmidos para transfección en células HeLa: Se construyeron los plásmidos LCR18-EGFP y LCR18-TELO-EGFP a partir de los plásmidos pEGFP-1 (Clonejet), que es un plásmido de expresión en mamíferos que funciona como un reportero EGFP y es útil para monitorear la transcripción de elementos regulatorios. La secuencia de la LCR tipo 18 fue obtenida por PCR a partir del plásmido LCR18-pGL3 basic y posteriormente clonada en el plásmido vector pJET1.2/blunt Cloning vector (Thermo, #1232) de acuerdo a procesos de ligación establecidos por el fabricante. Una vez ligado el inserto, el plásmido pJET1.2/blunt Cloning vector y pEGFP-1 fueron digeridos con la enzima de restricción BamHI y se ligó el inserto LCR tipo 18 con el plásmido pEGFP1 digerido de acuerdo con reacciones de ligación

estándar. Brevemente el inserto y el vector fueron ligados a una relación de (3:1) en un volumen de reacción de 20uL a 22 C toda la noche. Posteriormente se transformaron 5uL de esta reacción en células DH5alpha y con las colonias positivas se realizó PCR punto final para comprobar la orientación del inserto utilizando un primer forward dentro de la LCR y un primer reverse que alineara en la secuencia del gen reportero EGFP.

Cultivo celular: La línea celular HeLa, que corresponde a una línea celular derivada de un adenocarcinoma positivo a infección por HPV tipo 18, se mantuvo en medio DMEM suplementado al 10% con suero fetal bovino (Gibco) en una atmósfera de 5% de CO₂, 95% Aire a 37° C.

Tratamiento y selección con antibiótico G418: Se obtuvieron clones estables de la línea celular HeLa posterior a la transfección con los plásmidos LCR18-EGFP y LCR18-EGFP-Telo mediante la adición del antibiótico G418 (Geneticin, Gibco #11811031) a una concentración de 3 mg/mL, concentración a la cual todas las células mueren al tercer día obtenida mediante la estandarización de una curva de concentraciones desde 0.5 mg/mL hasta 4 mg/mL, en medio DMEM suplementado al 10% con SFB (Gibco).

Obtención de clones estables: Con el fin de obtener clones de las células transfectadas con los plásmidos anteriormente mencionados, se resuspendieron 50 células en un volumen de 10 mL de medio DMEM suplementado con SFB (10%). Una vez resuspendidas, se depositaron 100 uL de esta suspensión en una caja de 96 pozos. Con ayuda del microscopio, se verificó cuales pozos contenían una sola célula. Estas clones fueron posteriormente cultivadas para su posterior análisis.

Ensayos de Citometría de Flujo:

La línea celular HeLa será sembrada en placas de 6 pozos a una confluencia del 70% en medio DMEM (SFB 10%, 5% CO₂). Una vez alcanzada esta confluencia, será transfectada con las construcciones que contengan a la LCR unida al gen reportero EGFP y a una secuencia telomérica de acuerdo con las especificaciones del fabricante (Xfect, Takara). Después de 48h, estas células serán seleccionadas con el antibiótico G418 durante 4 semanas. Una vez transcurrido dicho tiempo, se obtendrán clones de estas poblaciones mediante dilución limitante en placas de 96 pozos. Luego, estas clones serán crecidas en medio DMEM (SFB 10%, CO₂ 5%) y serán evaluadas mediante ensayos de citometría de

flujo para observar la actividad del transgén EGFP durante un mes. Finalmente, las células serán cultivadas hasta alcanzar un 70% de confluencia y después serán levantadas con verseno, centrifugadas y resuspendidas en PBS (con paraformaldehído al 1%) para ser procesadas ese mismo día en el citómetro de flujo.

13. REFERENCIAS

1. Bouvard, V. et al. A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* 10, 321–322 (2009).
2. Jang, M. K., Shen, K. & McBride, A. A. Papillomavirus Genomes Associate with BRD4 to Replicate at Fragile Sites in the Host Genome. *PLoS Pathog.* 10, e1004117 (2014).
3. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* 69, 7–34 (2019).
4. McBride, A. A. Oncogenic human papillomaviruses. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160273 (2017).
5. McBride, A. A. & Warburton, A. The role of integration in oncogenic progression of HPV-associated cancers. *PLOS Pathog.* 13, e1006211 (2017).
6. Moody, C. A. & Laimins, L. A. Human papillomavirus oncoproteins: pathways to transformation. *Nat. Rev. Cancer* 10, 550–560 (2010).
7. zur Hausen, H. Papillomaviruses in the causation of human cancers — a brief historical account. *Virology* 384, 260–265 (2009).
8. Rigoni-Stern, D. Fatti statistici relative alle malattie cancerose che servirono de base alla poche cose dette dal dott. G. Serv. *Proprpatholterapser* 507–517 (1842).
9. de Villiers, E. M., Gissmann, L. & zur Hausen, H. Molecular cloning of viral DNA from human genital warts. *J. Virol.* 40, 932–935 (1981).
10. Hausen, H. zur. Papillomavirus infections — a major cause of human cancers. *Biochim. Biophys. Acta BBA - Rev. Cancer* 1288, F55–F78 (1996).
11. Hausen, H. HUMAN GENITAL CANCER: SYNERGISM BETWEEN TWO VIRUS INFECTIONS OR SYNERGISM BETWEEN A VIRUS INFECTION AND INITIATING EVENTS? *The Lancet* 320, 1370–1372 (1982).
12. Jablonska, S., Dabrowski, J. & Jakubowicz, K. Epidermodysplasia Verruciformis as a Model in Studies on the Role of Papovaviruses in Oncogenesis. *Cancer Res.* 32, 583–589 (1972).
13. Orth, G., Jablonska, S., Breitburd, F., Favre, M. & Croissant, O. The human papillomaviruses. *Bull. Cancer (Paris)* 65, 151–164 (1978).
14. Lacey, C. J. N., Lowndes, C. M. & Shah, K. V. Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. *Vaccine* 24, S35–S41 (2006).
15. Bernard, H.-U. Regulatory elements in the viral genome. *Virology* 445, 197–204 (2013).
16. Reference, G. H. What is a genome? Genetics Home Reference <https://ghr.nlm.nih.gov/primer/hgp/genome>.

17. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb. Perspect. Biol.* 2, (2010).
18. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25, 2227–2241 (2011).
19. Sandman, K., Pereira, S. L. & Reeve, J. N. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell. Mol. Life Sci. CMLS* 54, 1350–1364 (1998).
20. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* 421, 448–453 (2003).
21. Lakadamyali, M. & Cosma, M. P. Visualizing the genome in high resolution challenges our textbook understanding. *Nat. Methods* 17, 371–379 (2020).
22. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* 421, 448–453 (2003).
23. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 15, 2343–2360 (2001).
24. Ahmad, K. & Henikoff, S. Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci.* 99, 16477–16484 (2002).
25. Ganesan, A. Epigenetics: the first 25 centuries. *Philos. Trans. R. Soc. B Biol. Sci.* 373, (2018).
26. Felsenfeld, G. A Brief History of Epigenetics. *Cold Spring Harb. Perspect. Biol.* 6, (2014).
27. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* 403, 41–45 (2000).
28. Swagatika, S. & Tomar, R. S. Chapter Eight - Modulation of Epigenetics by Environmental Toxic Molecules. in *Advances in Molecular Toxicology* (eds. Fishbein, J. C. & Heilman, J. M.) vol. 10 361–389 (Elsevier, 2016).
29. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.* 34, 265–288 (2018).
30. Oberdoerffer, P. & Sinclair, D. A. The role of nuclear architecture in genomic instability and ageing. *Nat. Rev. Mol. Cell Biol.* 8, 692–702 (2007).
31. Baur, J. A., Zou, Y., Shay, J. W. & Wright, W. E. Telomere Position Effect in Human Cells. *Science* 292, 2075–2077 (2001).
32. Blasco, M. A. The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.* 8, 299–309 (2007).
33. Forsyth, N., Wright, W. & Shay, J. Telomerase and differentiation in multicellular organisms: Turn it off, turn it on, and turn it off again. *Differ. Res. Biol. Divers.* 69, 188–97 (2002).

34. Skvortzov, D., Rubtsova, M., Zvereva, M., Kiselev, F. & Donzova, O. The Regulation of Telomerase in Oncogenesis. *Acta Naturae* 1, 51–67 (2009).
35. Udugama, M. et al. Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res.* 43, 10227–10237 (2015).
36. García-Cao, M., O’Sullivan, R., Peters, A. H. F. M., Jenuwein, T. & Blasco, M. A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet.* 36, 94–99 (2004).
37. McBride, A. A. & Warburton, A. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathog.* 13, (2017).
38. Schwarz, E. et al. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* 314, 111–114 (1985).
39. Systematic Review of Genomic Integration Sites of Human Papillomavirus Genomes in Epithelial Dysplasia and Invasive Cancer of the Female Lower Genital Tract | *Cancer Research*. <https://cancerres.aacrjournals.org/content/64/11/3878.long>.
40. Yu, L. et al. Expression of HPV oncogenes from a single integration site in cervical cancer cell lines. *bioRxiv* 2020.06.25.171959 (2020) doi:10.1101/2020.06.25.171959.
41. Shukla, S. et al. Physical state & copy number of high risk human papillomavirus type 16 DNA in progression of cervical cancer. *Indian J. Med. Res.* 139, 531–543 (2014).
42. Jeon, S., Allen-Hoffmann, B. L. & Lambert, P. F. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J. Virol.* 69, 2989–2997 (1995).
43. Human Papillomavirus Type 16 Integrations in Cervical Tumors Frequently Occur in Common Fragile Sites | *Cancer Research*. <https://cancerres.aacrjournals.org/content/60/21/5916.long>.
44. Christiansen, I. K., Sandve, G. K., Schmitz, M., Dürst, M. & Hovig, E. Transcriptionally active regions are the preferred targets for chromosomal HPV integration in cervical carcinogenesis. *PLoS One* 10, e0119566 (2015).
45. Ferber, M. J. et al. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. *Oncogene* 22, 7233–7242 (2003).
46. Gao, G. et al. Common fragile sites (CFS) and extremely large CFS genes are targets for human papillomavirus integrations and chromosome rearrangements in oropharyngeal squamous cell carcinoma. *Genes. Chromosomes Cancer* 56, 59–74 (2017).
47. Glover, T. W., Arlt, M. F., Casper, A. M. & Durkin, S. G. Mechanisms of common fragile site instability. *Hum. Mol. Genet.* 14, R197–R205 (2005).
48. Ma, K. et al. Common Fragile Sites: Genomic Hotspots of DNA Damage and Carcinogenesis. *Int. J. Mol. Sci.* 13, 11974–11999 (2012).

49. Hecht, F. & Glover, T. W. Cancer chromosome breakpoints and common fragile sites induced by aphidicolin. *Cancer Genet. Cytogenet.* 13, 185–188 (1984).
50. Gao, G. & Smith, D. I. Role of the Common Fragile Sites in Cancers with a Human Papillomavirus Etiology. *Cytogenet. Genome Res.* 150, 217–226 (2016).
51. Jang, M. K., Shen, K. & McBride, A. A. Papillomavirus Genomes Associate with BRD4 to Replicate at Fragile Sites in the Host Genome. *PLoS Pathog.* 10, e1004117 (2014).
52. Nilsson, K., Wu, C. & Schwartz, S. Role of the DNA Damage Response in Human Papillomavirus RNA Splicing and Polyadenylation. *Int. J. Mol. Sci.* 19, 1735 (2018).
53. Sfeir, A. et al. Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. *Cell* 138, 90–103 (2009).
54. Morrish, T. A. et al. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446, 208–212 (2007).
55. Nambaru, L., Meenakumari, B., Swaminathan, R. & Rajkumar, T. Prognostic significance of HPV physical status and integration sites in cervical cancer. *Asian Pac. J. Cancer Prev. APJCP* 10, 355–360 (2009).
56. Kraus, I. et al. The Majority of Viral-Cellular Fusion Transcripts in Cervical Carcinomas Cotranscribe Cellular Sequences of Known or Predicted Genes. *Cancer Res.* 68, 2514–2522 (2008).
57. Ojesina, A. I. et al. Landscape of genomic alterations in cervical carcinomas. *Nature* 506, 371–375 (2014).
58. Rincón-Arano, H., Furlan-Magaril, M. & Recillas-Targa, F. Protection against telomeric position effects by the chicken cHS4 β -globin insulator. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14044–14049 (2007).
59. Hu, Z. et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* 47, 158–163 (2015).
60. Hu, Z. et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* 47, 158–163 (2015).
61. Gloss, B. & Bernard, H. U. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *J. Virol.* 64, 5577–5584 (1990).
62. Mohr, I. J. et al. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* 250, 1694–1699 (1990).
63. Cripe, T. P. et al. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and

repressor gene products: implications for cervical carcinogenesis. *EMBO J.* 6, 3745–3753 (1987).

64. Gloss, B., Bernard, H. U., Seedorf, K. & Klock, G. The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones. *EMBO J.* 6, 3735–3743 (1987).

65. Chong, T., Apt, D., Gloss, B., Isa, M. & Bernard, H. U. The enhancer of human papillomavirus type 16: binding sites for the ubiquitous transcription factors oct-1, NFA, TEF-2, NF1, and AP-1 participate in epithelial cell-specific transcription. *J. Virol.* 65, 5933–5943 (1991).

66. Mack, D. H. & Laimins, L. A. A keratinocyte-specific transcription factor, KRF-1, interacts with AP-1 to activate expression of human papillomavirus type 18 in squamous epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 88, 9102–9106 (1991).

67. Sv, G. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol.* 5, 1493–1506 (2010).

68. Thierry, F. Control of Human Papillomavirus type 18 transcription: role in carcinogenesis. 369–383 (1999).

69. Saavedra, A. L. La historia que no termina. 25 (2006).

70. Kadonaga, J. T. Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. *Cell* 116, 247–257 (2004).

71. Butler, J. E. F. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16, 2583–2592 (2002).

72. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* 13, 233–245 (2012).

73. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 144, 327–339 (2011).

74. Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* 5, 1 (2012).

75. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703–713 (2006).

76. Hark, A. T. et al. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489 (2000).

77. Braccioli, L. & de Wit, E. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* 63, 157–165 (2019).

78. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17507–17512 (2012).
79. Klenova, E. M. et al. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* 13, 7612–7624 (1993).
80. Kim, T. H. et al. Analysis of the vertebrate insulator protein CTCF binding sites in the human genome. *Cell* 128, 1231–1245 (2007).
81. Nakahashi, H. et al. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep.* 3, 1678–1689 (2013).
82. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein–DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* 36, 5221–5231 (2008).
83. Chen, H., Tian, Y., Shu, W., Bo, X. & Wang, S. Comprehensive Identification and Annotation of Cell Type-Specific and Ubiquitous CTCF-Binding Sites in the Human Genome. *PLoS ONE* 7, (2012).
84. Pentland, I. & Parish, J. L. Targeting CTCF to Control Virus Gene Expression: A Common Theme amongst Diverse DNA Viruses. *Viruses* 7, 3574–3585 (2015).
85. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012).
86. Van Bortle, K. et al. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* 15, R82 (2014).
87. Nora, E. P. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22 (2017).
88. 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).
89. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800 (2018).
90. Stigler, J., Çamdere, G. Ö., Koshland, D. E. & Greene, E. C. Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. *Cell Rep.* 15, 988–998 (2016).
91. Davidson, I. F. et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* 35, 2671–2685 (2016).
92. Nagy, G. et al. Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genomics* 17, 637 (2016).
93. Rao, S. S. P. et al. Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320.e24 (2017).

94. Pentland, I. et al. Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol.* 16, e2005752 (2018).
95. Burley, M., Roberts, S. & Parish, J. L. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin. Immunopathol.* 42, 159–171 (2020).
96. Atchison, M. L., Basu, A., Zaprazna, K. & Papasani, M. Mechanisms of Yin Yang 1 in Oncogenesis: The Importance of Indirect Effects. *Crit. Rev. Oncog.* 16, (2011).
97. Katainen, R. et al. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* 47, 818–821 (2015).
98. CTCFBSDB 2.0: A database for CTCF binding sites and genome organization. <http://insulatordb.uthsc.edu/>.
99. Schmidt, D. et al. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell* 148, 335–348 (2012).
100. HeLa ATCC® CCL-2TM. <https://www.atcc.org/en/Products/All/CCL-2.aspx>.
101. Hall, A. H. S. & Alexander, K. A. RNA Interference of Human Papillomavirus Type 18 E6 and E7 Induces Senescence in HeLa Cells. *J. Virol.* 77, 6066–6069 (2003).
102. Martínez-Ramírez, I. et al. SOX2 as a New Regulator of HPV16 Transcription. *Viruses* 9, (2017).
103. de Lange, T. Human telomeres are attached to the nuclear matrix. *EMBO J.* 11, 717–724 (1992).
104. Soto-Reyes, E. & Recillas-Targa, F. Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene* 29, 2217–2227 (2010).
105. Adey, A. et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* 500, 207–211 (2013).
106. You, J. Papillomavirus interaction with cellular chromatin. *Biochim. Biophys. Acta* 1799, 192–199 (2010).
107. Paris, C. et al. CCCTC-Binding Factor Recruitment to the Early Region of the Human Papillomavirus 18 Genome Regulates Viral Oncogene Expression. *J. Virol.* 89, 4770–4785 (2015).
108. Pentland, I. et al. Disruption of CTCF-YY1-dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLoS Biol.* 16, e2005752 (2018).
109. Burley, M., Roberts, S. & Parish, J. L. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Semin. Immunopathol.* 42, 159–171 (2020).

110. Hart, K. W. et al. Novel Method for Detection, Typing, and Quantification of Human Papillomaviruses in Clinical Samples. *J. Clin. Microbiol.* 39, 3204–3212 (2001).
111. Phillips-Cremins, J. E. & Corces, V. G. Chromatin Insulators: Linking genome organization to cellular function. *Mol. Cell* 50, 461–474 (2013).
112. Akagi, K. et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* 24, 185–199 (2014).
113. Van Doorslaer, K. et al. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res.* 45, D499–D506 (2017).
114. Wang, X., Cairns, M. J. & Yan, J. Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res.* 47, 11481–11496 (2019).
115. Shen, C. et al. Long-distance interaction of the integrated HPV fragment with MYC gene and 8q24.22 region upregulating the allele-specific MYC expression in HeLa cells. *Int. J. Cancer* 141, 540–548 (2017).
116. Williams, V. M., Filippova, M., Soto, U. & Duerksen-Hughes, P. J. HPV-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress. *Future Virol.* 6, 45–57 (2011).
117. Rincón-Arano, H., Furlan-Magaril, M. & Recillas-Targa, F. Protection against telomeric position effects by the chicken cHS4 β -globin insulator. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14044–14049 (2007).
118. Pedram, M. et al. Telomere Position Effect and Silencing of Transgenes near Telomeres in the Mouse. *Mol. Cell. Biol.* 26, 1865–1878 (2006).
119. Park, I.-S. et al. Characterization of the methylation patterns in human papillomavirus type 16 viral DNA in head and neck cancers. *Cancer Prev. Res. Phila. Pa* 4, 207–217 (2011).
120. Kallestad, L., Christensen, K., Woods, E. & Milavetz, B. Transcriptional repression is epigenetically marked by H3K9 methylation during SV40 replication. *Clin. Epigenetics* 6, (2014).
121. Wang, W. et al. Expression of Yin Yang 1 in cervical cancer and its correlation with E-cadherin expression and HPV16 E6. *PloS One* 13, e0193340 (2018).
122. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *EMBO Rep.* 16, 1467–1481 (2015).