

Universidad de Salamanca
Departamento de Microbiología y Genética

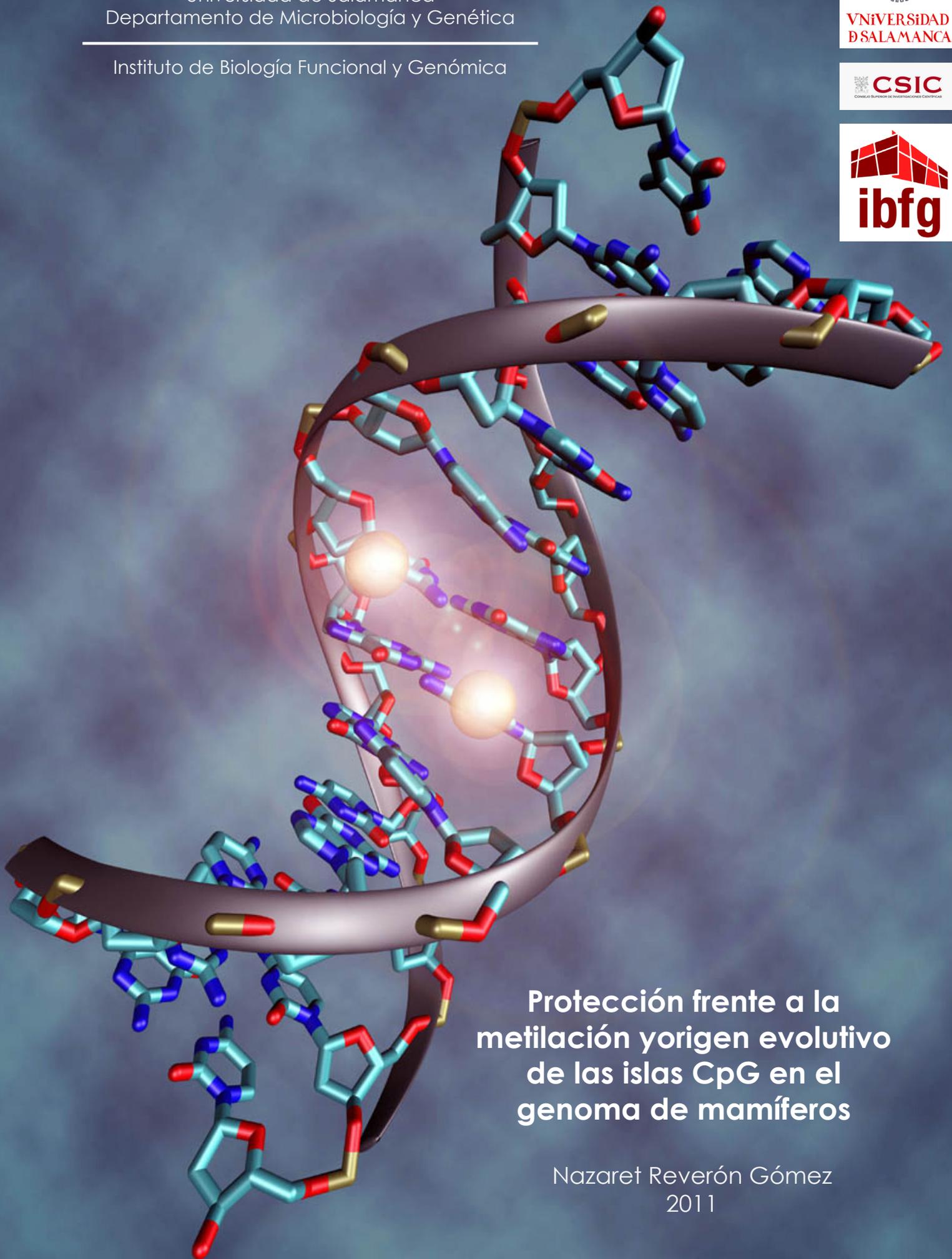
Instituto de Biología Funcional y Genómica



UNIVERSIDAD
DE SALAMANCA

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

ibfg



**Protección frente a la
metilación y origen evolutivo
de las islas CpG en el
genoma de mamíferos**

Nazaret Reverón Gómez
2011



VNiVERSiDAD
D SALAMANCA



INSTITUTO DE BIOLOGÍA
FUNCIONAL Y GENÓMICA

Universidad de Salamanca
Departamento de Microbiología y Genética

Instituto de Biología Funcional y Genómica

PROTECCIÓN FRENTE A LA METILACIÓN Y ORIGEN EVOLUTIVO DE LAS ISLAS CpG EN EL GENOMA DE MAMÍFEROS

Nazaret Reverón Gómez
2011



**VNiVERSiDAD
D SALAMANCA**



Dr. Ángel Domínguez Olavarri, Director del Departamento de Microbiología y Genética de la Universidad de Salamanca

CERTIFICA:

Que la memoria titulada "Protección frente a la metilación y origen evolutivo de las islas CpG en el genoma de mamíferos" presentada por la licenciada Dña. Nazaret Reverón Gómez para optar al grado de Doctor en Biología por la Universidad de Salamanca, ha sido realizada bajo la dirección del Dr. Francisco Antequera Márquez en el Centro Mixto Instituto de Biología Funcional y Genómica, CSIC/Departamento de Microbiología y Genética de la Universidad de Salamanca. Y para que así conste, expide el siguiente certificado en Salamanca, a 25 de Mayo de 2011.

Fdo. Ángel Domínguez Olavarri

Dr. Francisco Antequera Márquez, Profesor de Investigación del Consejo Superior de Investigaciones Científicas

CERTIFICA:

Que la memoria titulada "Protección frente a la metilación y origen evolutivo de las islas CpG en el genoma de mamíferos" presentada por la licenciada Dña. Nazaret Reverón Gómez para optar al grado de Doctor en Biología por la Universidad de Salamanca, ha sido realizada bajo su dirección en el Centro Mixto Instituto de Biología Funcional y Genómica, CSIC/Departamento de Microbiología y Genética de la Universidad de Salamanca. Y para autorizar su presentación y evaluación por el tribunal correspondiente, firma el siguiente certificado en Salamanca, a 25 de Mayo de 2011.

Fdo. Francisco Antequera Márquez



**VNIVERSIDAD
D SALAMANCA**



Dra. Yolanda Sánchez Martín, Profesora Titular del Departamento de Microbiología y Genética de la Universidad de Salamanca

CERTIFICA:

Que la memoria titulada "Protección frente a la metilación y origen evolutivo de las islas CpG en el genoma de mamíferos" presentada por la licenciada Dña. Nazaret Reverón Gómez para optar al grado de Doctor en Biología por la Universidad de Salamanca, ha sido realizada bajo la dirección del Dr. Francisco Antequera Márquez en el Centro Mixto Instituto de Biología Funcional y Genómica, CSIC/Departamento de Microbiología y Genética de la Universidad de Salamanca. Y para que así conste, expide el siguiente certificado en Salamanca, a 25 de Mayo de 2011.

Fdo. Yolanda Sánchez Martín

Esta tesis doctoral ha sido financiada por una beca predoctoral del Programa de Formación del Profesorado Universitario (2006/2010) del Ministerio de Educación del Gobierno de España, proyectos del Plan Nacional de I+D y por el proyecto CONSOLIDER Ingenio 2010.

“La verdadera ciencia enseña, por encima de todo, a dudar y a ser ignorante.”

Miguel de Unamuno

AGRADECIMIENTOS

En primer lugar gracias a Paco por la oportunidad que me ha brindado, así como por la confianza ciega que ha depositado en mí a lo largo de estos años y por haber logrado transmitirme su pasión por la ciencia.

A Mariano, el mejor docente que he tenido el placer de conocer, y el principal responsable de que me dedique a la investigación. Gracias por el apoyo prestado durante estos años y por el cariño con el que me recibes cada Navidad en las Jornadas de “El Almendro”.

A María Gómez, por su paciencia, su apoyo, su interés y por haberme enseñado tantas cosas sin tener por qué hacerlo.

A Laura de los Ríos, Cristina, Elisa y Laura Mojardín, por acogerme con cariño cuando llegué al frío charro desde tierras más cálidas. A Laura de los Ríos, por enseñarme todo lo que sé de bisulfito, así como muchos aspectos prácticos de la vida que tan útiles me han resultado. A Laura Mojardín, por su humor, su afán por hacer las cosas bien hechas y por el mejor arroz con leche del mundo que tantos momentos dulces nos regaló. A Cristina y Elisa por su apoyo, su cariño y sus buenos consejos.

A Laura Marín por su cariño, su amistad y su inestimable apoyo en momentos de agobio. A Mar, por su alegría, su cercanía y por ser un ejemplo de fortaleza y superación. Y por supuesto por las fiestas en tu piscina, que tanto nos alegran los veranos. A Joana, de la que tantas cosas he aprendido, gracias por tu amistad, tu confianza y tu apoyo.

A los que llegaron después, Nacho y Rebeca, por escuchar pacientemente mis lamentos y aportarnos la ilusión y el entusiasmo de los que comienzan. A Jose, que durante estos últimos meses me ha acompañado en la soledad de las islas CpG. Gracias por tu paciencia y por la ayuda que me has prestado con la tesis.

A todos los becarios con los que he compartido tantos buenos momentos durante estos años, en especial a mis compañeros de escritura, Talía, Esther, Javi y Alberto, por los momentos de desahogo tan necesarios en esta última etapa.

A todos los investigadores del IBFG, por su interés, su ayuda y sus consejos.

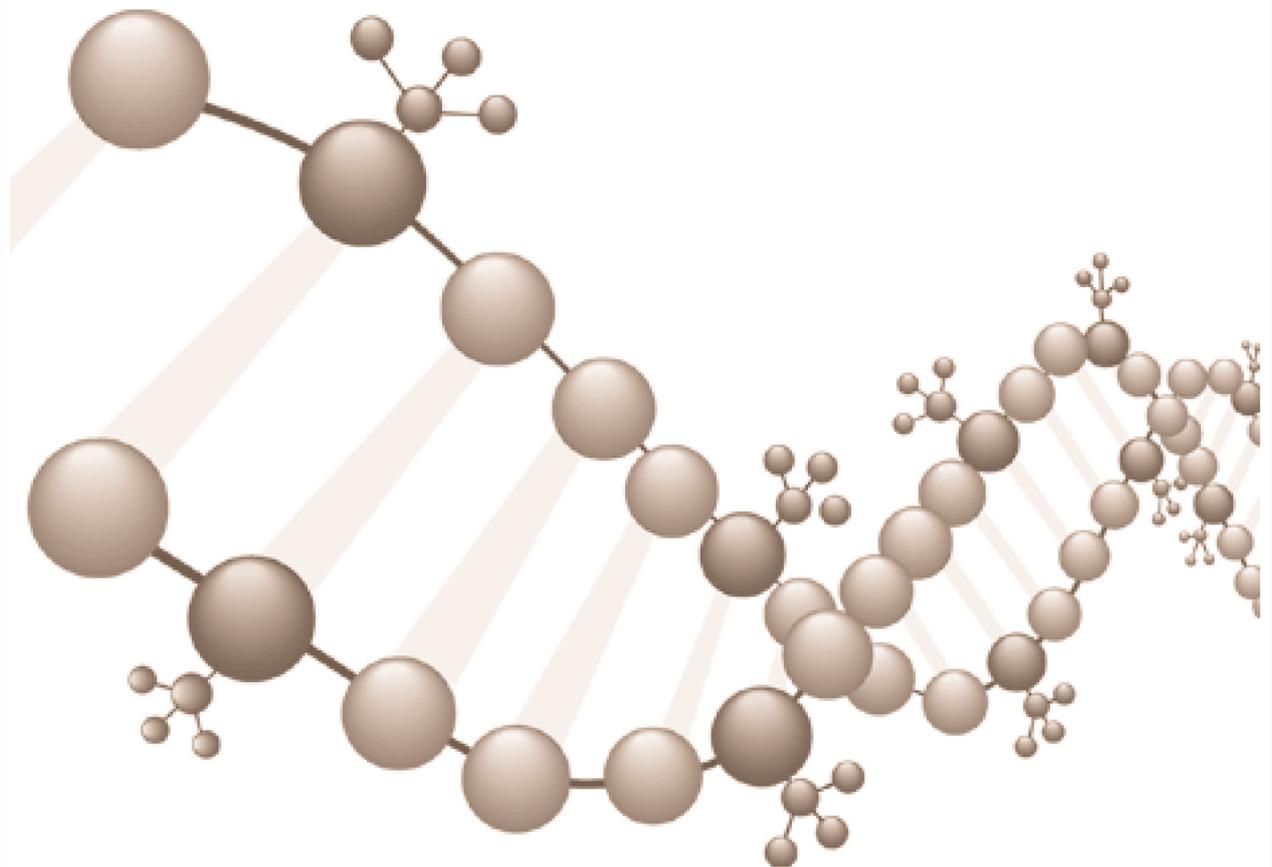
A todo el personal técnico y administrativo del IBFG, por su amabilidad y su disposición a ayudar. En especial a Paco Alonso, por arreglar todo lo arreglable sin perder nunca la sonrisa y a Paco Soriano por ponerle siempre un punto de humor a los viajes por el pasillo.

A mis amigos, los de aquí y los de allá, que son pocos pero irremplazables. A mi gente de Tenerife, porque citando a otro chicharrero *“Mi casa está en el mar con siete puertas, yo ya no vivo allí pero me esperan”*. Y por esperarme siempre con los brazos abiertos tengo que darles las gracias a todos. A mi Leti, la primera y mejor amiga que he tenido. A Sandra, porque todavía hoy me sigues abriendo la mente. A Cecilia, por tu alegría y tu cariño. A Juana y Virginia, porque da gusto sentir al hablar con ustedes que no importa el tiempo que pase ni la distancia que nos separe. Gracias a Juana y Ale, por seguirme en mi aventura Salmantina. Estos años en la estepa charra no hubieran sido lo mismo sin ustedes, no saben cuánto les echo de menos. A mis amigos de Cional, con los que he crecido verano a verano y a los que considero prácticamente como de mi familia (a los que no lo son, claro). En especial a Elisabet, Carlos, Ruth, David y Noa, porque siempre se puede contar con vosotros.

A mi familia, la biológica y la política, por todo el amor, la confianza y el apoyo, y sobre todo a mis padres, por una vida de sacrificios para que nunca me faltara de nada. Espero no decepcionaros nunca.

Y por supuesto, a Álvaro, por quererme, cuidarme, soportarme y por estar dispuesto a seguirme al fin del mundo sin exigir nada a cambio. Mi vida no estaría completa sin ti.

ÍNDICE

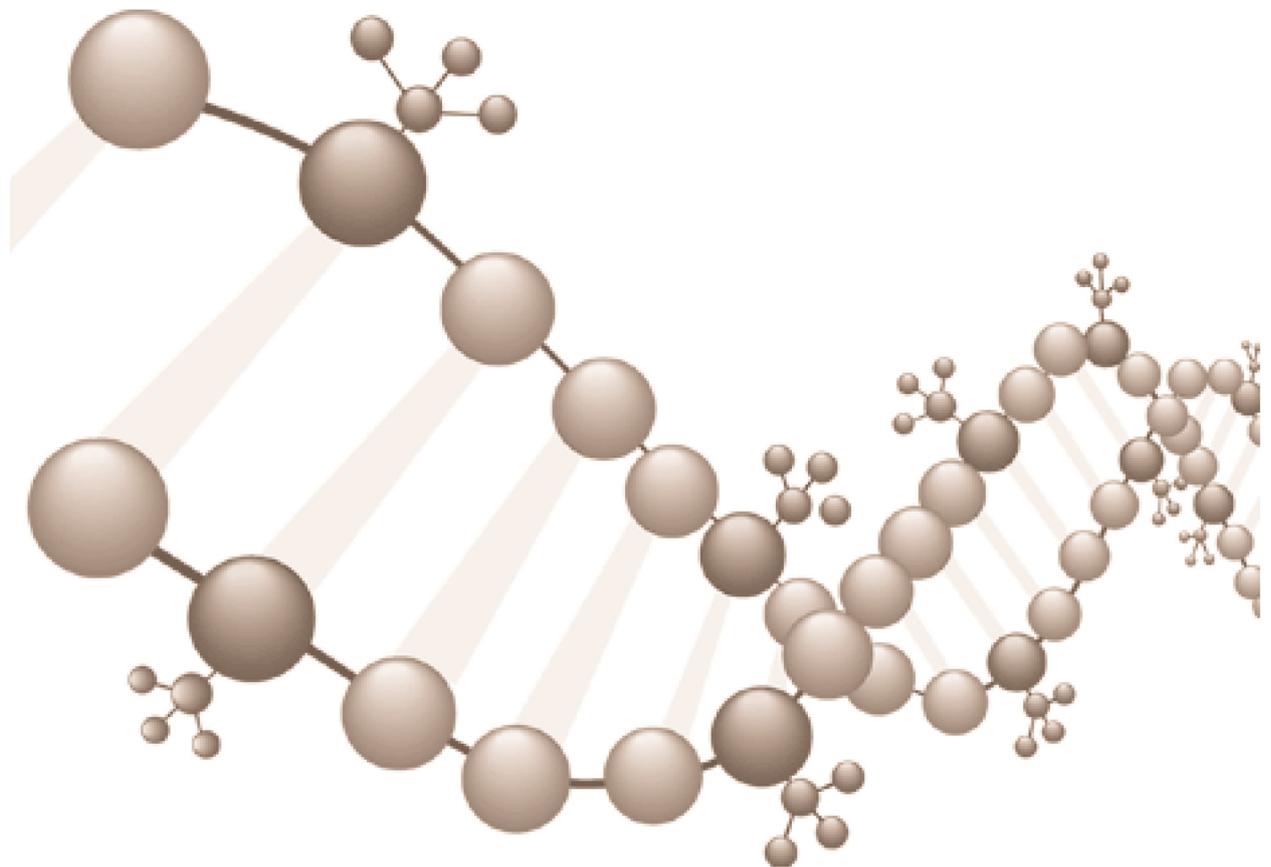


INTRODUCCIÓN	9
1. METILACIÓN DEL DNA	9
1.1. Metilación del DNA en el genoma de eucariotas	9
1.2. DNA metil-transferasas	10
1.3. Proteínas de unión a DNA metilado	11
1.4. Establecimiento y mantenimiento de los patrones de metilación en el genoma de mamíferos	13
1.5. Funciones de la metilación genómica en mamíferos	15
1.6. Distribución de la metilación y organización de los dinucleótidos CpG en el genoma de mamíferos	17
2. ISLAS CpG	18
2.1. Propiedades de las islas CpG	18
2.2. Islas CpG y promotores	19
2.3. Islas CpG y orígenes de replicación	21
2.4. Islas CpG y metilación	22
OBJETIVOS	29
RESULTADOS	33
1. PAPEL DE LA REPLICACIÓN DEL DNA EN LA PROTECCIÓN FRENTE A LA METILACIÓN DE LAS ISLAS CpG	33
1.1. Antecedentes	33
1.2. Conservación filogenética de los ORIs de mamíferos que se encuentran asociados a islas CpG	34
1.3. Análisis de metilación de los ORIs LMNB2 y CDKN2B	39
1.4. La unión de ORC como barrera frente a la metilación en el extremo 5' de las islas CpG	40
2. CONTRIBUCIÓN DE LA TRANSCRIPCIÓN AL MANTENIMIENTO DE LAS ISLAS CpG EN ESTADO NO METILADO	51
2.1. Antecedentes	51
2.2. Análisis de metilación de islas asociadas a genes específicos de tejido en ausencia de transcripción activa	51
2.3. Análisis de metilación de la isla asociada al gen de la α -globina en condiciones de transcripción activa	57
2.4. Análisis del patrón de metilación de la isla CpG asociada al gen de la α -globina en ausencia de transcripción activa	59
2.5. Análisis de metilación mediante MeDIP combinado con la hibridación de <i>arrays</i> de alta densidad (MeDIP-CHIP)	60
2.5.1. Diseño experimental	60
2.5.2. Metilación de DNA <i>in vitro</i>	62
2.5.3. MeDIP-PCR	62

2.5.4. Amplificación de DNA inmunoprecipitado	63
2.5.5. Hibridación de <i>tiling microarrays</i>	64
2.5.6. Análisis de los datos de MeDIP-CHIP	64
2.5.7. Validación de los resultados por MeDIP-PCR y análisis de metilación por bisulfito sódico	65
2.5.8. Análisis comparativo de islas CpG con distintos niveles de metilación	68
3. DINÁMICA DE GENERACIÓN DE LAS ISLAS CpG EN EL GENOMA DE MAMÍFEROS	71
3.1. Antecedentes	71
3.2. Identificación y caracterización de regiones no metiladas del genoma que no se clasifican como islas CpG (MeDIP-CHIP)	72
3.3. Identificación y análisis de una isla CpG en proceso de formación en el genoma de <i>Mus musculus</i>	77
DISCUSIÓN	87
1. PAPEL DE LA REPLICACIÓN DEL DNA EN LA PROTECCIÓN FRENTE A LA METILACIÓN DE LAS ISLAS CpG	87
2. CONTRIBUCIÓN DE LA TRANSCRIPCIÓN AL MANTENIMIENTO DE LAS ISLAS CpG EN ESTADO NO METILADO	91
2.1. Metilación diferencial de islas CpG y regulación de la expresión génica	91
2.2. Contribución de la transcripción activa a la protección de las islas CpG frente a la metilación	92
3. DINÁMICA DE GENERACIÓN DE LAS ISLAS CpG EN EL GENOMA DE MAMÍFEROS	97
3.1. Inestabilidad genética en los promotores asociados a islas CpG	97
3.2. Incremento en el contenido en G+C en las regiones inestables del genoma	99
3.3. Dinámica de generación de las islas CpG en el genoma de mamíferos y mantenimiento del estado no metilado	102
CONCLUSIONES	109
MATERIALES Y MÉTODOS	113
1. CULTIVOS CELULARES	113
2. PURIFICACIÓN DE ÁCIDOS NUCLÉICOS	113
2.1. Extracción de DNA a partir de células en cultivo	113
2.2. Extracción de DNA de sangre	113
2.3. Extracción de DNA de esperma	114
2.4. Purificación de cadenas nacientes de DNA	114
3. TÉCNICAS DE ANÁLISIS DE METILACIÓN DE DNA	115
3.1. Análisis de metilación con bisulfito sódico	115
3.1.1. Fragmentación del DNA genómico	116
3.1.2. Desnaturalización del DNA fragmentado	116

3.1.3. Transformación con bisulfito sódico	116
3.1.4. Análisis por PCR	117
3.1.5. Clonación y obtención de DNA plasmídico	118
3.2. Inmunoprecipitación de DNA metilado (MeDIP)	118
3.2.1. Sonicación del DNA genómico	118
3.2.2. Desnaturalización	118
3.2.3. Inmunoprecipitación de DNA metilado	118
4. INMUNOPRECIPITACIÓN DE CROMATINA	120
4.1. <i>Cross-linking</i>	120
4.2. Extracción de los núcleos y eliminación del material citosólico	120
4.3. Lisis y fragmentación de la cromatina	120
4.4. Inmunoprecipitación	121
4.5. Reversión del <i>cross-linking</i> y purificación del DNA inmunoprecipitado	121
5. PCR CUANTITATIVA	122
6. METILACIÓN DE DNA GENÓMICO "IN VITRO"	122
7. HIBRIDACIÓN Y ANÁLISIS DE <i>TILING MICROARRAYS</i>	123
7.1. Amplificación de DNA inmunoprecipitado	123
7.2. Fragmentación y marcaje del DNA amplificado	124
7.3. Hibridación de los <i>tiling microarrays</i>	124
7.4. Análisis de datos	124
8. HERRAMIENTAS BIOINFORMÁTICAS	124
8.1. Navegadores especializados	124
8.2. Programas de alineamientos múltiples	125
8.3. Editor de alineamientos	126
8.4. Bases de datos	126
8.4.1. TIGER	126
8.4.2. DBTSS	126
8.4.3. FANTOM	127
BIBLIOGRAFÍA	131
INFORMACIÓN SUPLEMENTARIA	145
ABREVIATURAS	161

INTRODUCCIÓN



1. METILACIÓN DEL DNA

1.1. Metilación del DNA en el genoma de eucariotas

La metilación del DNA es una modificación covalente presente en numerosos organismos, tanto procariotas como eucariotas, aunque los nucleótidos que se metilan, al igual que su distribución genómica, varían entre los distintos grupos. En procariotas, la adición de un grupo metilo puede ocurrir tanto en adeninas como en citosinas, mientras que en eucariotas la metilación se produce exclusivamente en posición 5' del anillo de pirimidina de las citosinas. Dentro de los eucariotas, la metilación de citosinas se produce en secuencias nucleotídicas distintas dependiendo del organismo (Feng *et al.*, 2010; Zemach *et al.*, 2010). Así, en animales la metilación se produce casi exclusivamente en citosinas seguidas de una guanina en la misma cadena, es decir, dinucleótidos CG o CpG, donde la "p" representa el enlace fosfato entre ambos nucleótidos. La metilación en CpGs es simétrica, es decir, tiene lugar en ambas cadenas del DNA. En plantas, sin embargo, la metilación suele ocurrir en citosinas que se encuentran incluidas en CG, CHG o CHH (H=A, T o C), de modo que puede ser simétrica en el caso de CG y CHG, o asimétrica en el caso de CHH. Por su parte, los hongos suelen presentar metilación en citosinas incluidas en dinucleótidos CG y CH. No obstante, y pese a la amplia distribución de esta modificación covalente entre los distintos taxones, no todos los eucariotas presentan metilación del DNA. Los genomas de organismos modelo como *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* o *Caenorhabditis elegans* carecen de metilación y en el caso de *Drosophila melanogaster* sólo se han detectado niveles muy bajos de metilación, la cual suele tener lugar mayoritariamente en los dinucleótidos CT o CA (Lyko *et al.*, 2000).

Los vertebrados poseen los mayores niveles de metilación en CG (Fig. 1) y ésta se distribuye de forma global por todo el genoma, a excepción de pequeñas regiones, generalmente asociadas a los promotores de los genes, que se denominan islas CpG. En el caso de plantas y hongos, así como en invertebrados como *Ciona intestinalis*, la metilación presenta generalmente una distribución en mosaico, es decir, existen dominios de DNA metilado intercalados con dominios de DNA no metilados (Simmen *et al.*, 1999; Suzuki & Bird, 2008).

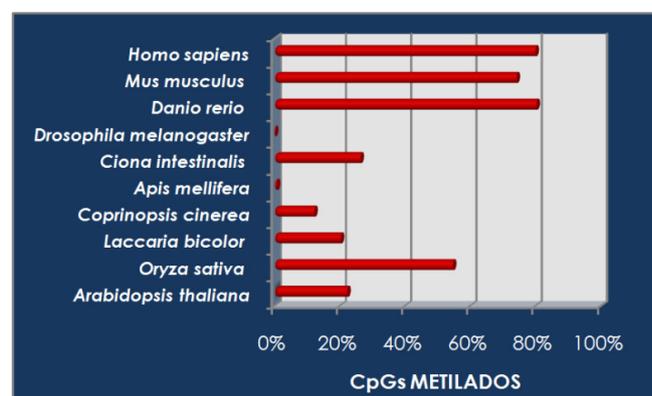


Figura 1. Niveles de metilación en dinucleótidos CpG en varios organismos eucariotas.

Recientemente se han llevado a cabo estudios de metilación a nivel genómico de varios organismos, entre los que se encuentran plantas como, *Oryza sativa* (arroz) o *Arabidopsis thaliana*, hongos como *Laccaria bicolor* y *Coprinopsis cinerea*, invertebrados como *Apis mellifera* (abeja) y *Ciona intestinalis* (ascidia) y vertebrados como *Danio rerio* (pez cebra) o *Mus musculus* (ratón). Estos análisis han permitido determinar que la metilación en el cuerpo de los genes se conserva en la mayoría de los organismos. Sin embargo, la metilación de los elementos transponibles, presente en plantas y animales, parece haber evolucionado de forma independiente en estos dos linajes, ya que no se detecta en ninguno de los invertebrados analizados (Feng *et al.*, 2010; Zemach *et al.*, 2010).

1.2. DNA metil-transferasas

Las proteínas responsables de la adición de un grupo metilo en posición 5' del anillo de pirimidina de las citosinas se denominan DNA metiltransferasas. Están presentes en numerosos organismos, tanto procariotas como eucariotas, y se caracterizan por compartir 10 motivos aminoacídicos, 6 de los cuales se encuentran muy conservados a lo largo de la evolución (Colot & Rossignol, 1999; Goll & Bestor, 2005). La mayoría de las DNA metiltransferasas se encuentran englobadas dentro de 4 familias: DNMT1, DNMT2, DNMT3 y cromometilasas. La familia de las cromometilasas es exclusiva del reino vegetal. Todos los organismos que poseen metiltransferasas de la familia DNMT1 incluyen también algún miembro de las familias DNMT2 y DNMT3, mientras que DNMT2 parece ser la única metiltransferasa en algunos organismos. El número de metiltransferasas varía ampliamente entre los distintos eucariotas, desde organismos como *D. melanogaster* que presenta a DNMT2 como única metiltransferasa hasta *Arabidopsis thaliana*, en la que se han descrito diez proteínas de este tipo. Por otra parte, organismos como *C. elegans* o *S. cerevisiae*, que carecen de metilación genómica, no presentan ninguna secuencia codificante homóloga a las metiltransferasas.

En mamíferos, concretamente en humano y ratón, se han descrito cuatro DNA metiltransferasas, que se clasifican en metiltransferasas *de novo* o de mantenimiento en función de que su sustrato sea DNA no metilado o DNA hemimetilado, respectivamente (**Tabla 1**). En la categoría de metiltransferasas *de novo* están DNMT3A y DNMT3B, que son las responsables de establecer los patrones de metilación durante el desarrollo embrionario temprano. Se expresan mayoritariamente en células ES (*Embryonic Stem Cells*: Células madre embrionarias), aunque también se encuentran en menor nivel en muchos tejidos somáticos. Se ha descrito una segunda isoforma de DNMT3A, DNMT3A2, que es específica de línea germinal masculina. Dentro de la familia DNMT3 se ha descrito otra proteína que presenta homología con DNMT3A y DNMT3B pero carece de actividad metiltransferasa. Se trata de DNMT3L, una proteína que se

expresa exclusivamente en línea germinal y fases tempranas del desarrollo embrionario y que es esencial para el establecimiento de los patrones de metilación durante la gametogénesis (Bourc'his *et al.*, 2001). DNMT3L ejerce su función mediante la interacción directa de su dominio carboxi-terminal con DNMT3A y DNMT3B (Suetake *et al.*, 2004).

DNMT1 es la metiltransferasa de mantenimiento, es decir, la que se encarga de copiar los patrones de metilación de la hebra parental a la hebra recién sintetizada cuando tiene lugar la replicación del DNA. De acuerdo con esto, muestra sus mayores niveles de expresión durante la fase S del ciclo celular y posee dominios que interaccionan con PCNA, lo que asegura su localización en los focos de replicación y con UHRF1, proteína que se une específicamente a DNA hemimetilado y que parece ser la responsable del reclutamiento de DNMT1 (Law & Jacobsen, 2010). DNMT1 presenta una segunda isoforma, DNMT1o, que se expresa específicamente en el oocito y ha sido implicada en el mantenimiento de la metilación de los genes de *imprinting* durante la reprogramación epigenética que tiene lugar en el desarrollo embrionario temprano.

Existe además una cuarta DNA metiltransferasa en mamíferos que no encaja en ninguna de las dos categorías anteriores. Se trata de DNMT2, una proteína de expresión ubicua que, a pesar de poseer homología de secuencia y estructura con el resto de DNA metiltransferasas no muestra actividad DNA metiltransferasa *in vitro* y la delección del gen no causa ningún fenotipo apreciable, ni siquiera en organismos como *D. melanogaster* donde es la única DNA metiltransferasa presente (Goll & Bestor, 2005). En 2006 Goll *et al.* demostraron que DNMT2 es capaz de metilar RNA en humanos, en concreto la citosina 38 del RNA de transferencia del ácido aspártico (tRNA-Asp), y que esta función se conserva en ratón, *Drosophila melanogaster* y *Arabidopsis thaliana*. Recientemente se ha visto que además es capaz de metilar otros tRNAs (tRNA-Val y tRNA-Gly) y que esta metilación podría constituir un mecanismo de defensa contra la degradación por ribonucleasas en condiciones de estrés (Schaefer *et al.*, 2010).

Tabla 1. DNA metiltransferasas

DNA METILTRANSFERASAS	
DNMT1	Mantenimiento de los patrones de metilación
DNMT2	Metilación de RNAs de transferencia
DNMT3 A/B	Establecimiento de los patrones de metilación en el embrión y la línea germinal
DNMTL	Factor regulador que recluta a DNMT3A/B al DNA

1.3. Proteínas de unión a DNA metilado

En mamíferos se han descrito tres familias de proteínas de unión a DNA metilado: las proteínas MBD, las proteínas SAR y las proteínas de unión al DNA metilado por dedos de zinc

(Sasai & Defossez, 2009) (**Tabla 2**). Las proteínas MBD fueron las primeras en describirse y se caracterizan porque contienen un dominio común de unión a DNA metilado (MBD) capaz de reconocer específicamente CpGs metilados. Dentro de esta familia encontramos cinco proteínas, MeCP2, MBD1, MBD2, MBD3 y MBD4 (Klose & Bird, 2006).

MeCP2, MBD1 y MBD2 funcionan como represores transcripcionales reclutando diversos complejos entre los que se incluyen desacetilasas y metiltransferasas de histonas y otros factores remodeladores de la cromatina. Así por ejemplo MeCP2 interacciona, entre otros, con el complejo SIN3A/HDAC, que incluye represores transcripcionales y desacetilasas de histonas. MBD2, por su parte, recluta el complejo represor NuRD/Mi-2 y MBD1 se asocia a las metiltransferasas de histonas SETDB1 y SUV39H1, así como a la proteína de unión a heterocromatina HP-1.

MBD3 forma parte del complejo represor NuRD que es reclutado por MBD2, pero carece de la capacidad de unirse específicamente al DNA metilado a pesar de su elevada homología con las otras proteínas de esta familia. MBD4, por su parte, sí se une específicamente a DNA metilado pero su función principal es la de reparación de emparejamientos erróneos T:G generados como consecuencia de la desaminación espontánea de las citosinas metiladas mediante un dominio con actividad glicosilasa (Millar *et al.*, 2002). No obstante, se ha observado que también puede ejercer represión transcripcional mediante su interacción con el complejo Sin3A/HDAC (Kondo *et al.*, 2005).

Tabla2. Proteínas de unión a DNA metilado

PROTEÍNAS DE UNIÓN A DNA METILADO	
MeCP1	Represión transcripcional a través de la interacción con complejos remodeladores de la cromatina
MBD1	
MBD2	
MBD3	Componente del complejo represor NuRD, reclutado por MBD2
MBD4	Participa en la reparación de emparejamientos erróneos T:G
UHRF1	Recluta a DNMT1 al DNA hemimetilado durante la fase S del ciclo celular
UHRF2	Función desconocida
KAISO	Represión transcripcional a través del reclutamiento del complejo N-CoR
ZBTB4	
ZBTB38	Función desconocida

La familia de proteínas SAR incluye a UHRF1 y UHRF2 y se caracteriza por un dominio SAR (*SET and Ring- associated*) que es capaz de unirse específicamente al DNA hemimetilado. Esta especificidad de unión junto con la interacción de UHF1 con la DNMT1 durante la fase S del ciclo celular confiere a esta proteína un papel esencial en el mantenimiento de los patrones de metilación del DNA (Sasai & Defossez, 2009).

Finalmente, dentro de la familia de proteínas de unión a DNA metilado por dedos de zinc se han descrito tres proteínas: KAISO, ZBTB4 y ZBTB38. Como su nombre indica poseen varios

dedos de zinc a través de los cuales reconocen el DNA metilado por un mecanismo que parece ser dependiente de secuencia (Sasai *et al.*, 2010). KAISO reprime la transcripción de determinados genes mediante el reclutamiento, a través de un dominio BTB, del complejo N-CoR, que presenta actividad desacetilasa de histonas. No existen evidencias acerca del papel de ZBTB4 y ZBTB38 como represores transcripcionales, pero también poseen un dominio BTB, lo que sugiere un papel similar al que desempeña KAISO.

1.4. Establecimiento y mantenimiento de los patrones de metilación en el genoma de mamíferos

El establecimiento de los patrones de metilación del DNA en mamíferos tiene lugar durante las etapas tempranas del desarrollo embrionario mediante dos procesos de reprogramación epigenética, uno en los precursores germinales durante la gametogénesis y otro a nivel del embrión justo después de la fertilización del óvulo (Fig. 2).

La primera reprogramación epigenética del genoma tiene lugar en los precursores germinales (PGCs) durante la gametogénesis. Tras la migración de los PGCs a las gónadas en desarrollo se produce en estas células una desmetilación global que provoca la eliminación de las marcas parentales específicas o marcas de *imprinting* y su restablecimiento *de novo* en función del sexo del individuo.

Los genes de *imprinting* son un pequeño conjunto de genes en los que se ha descrito transcripción monoalélica dependiendo del origen parental, es decir, se expresan sólo desde el alelo paterno o materno. Esta expresión monoalélica está sujeta a marcas parentales específicas de tipo epigenético que se establecen en la línea germinal y se mantienen durante el desarrollo del organismo. Estos genes suelen disponerse en *clusters* cuya expresión está regulada por elementos que actúan en *cis* denominados centros reguladores de *imprinting* (ICRs). Estos ICRs se encuentran diferencialmente metilados en función del sexo del individuo y esta metilación diferencial, junto con modificaciones de histonas específicas, constituyen la marca de *imprinting*, que debe ser reprogramada durante la gametogénesis para que los gametos maduros posean la marca específica del sexo del individuo en el que se generan.

Aún no ha sido determinado el mecanismo ni las enzimas implicadas en la desmetilación que tiene lugar en los PGCs, aunque recientemente se ha observado que está ligada a la aparición de roturas de cadena simple y la actividad de la maquinaria de reparación por escisión de base (BER: Base Excision Repair) (Hajkova *et al.*, 2010). Tras la desmetilación tiene lugar una oleada de metilación *de novo* llevada a cabo por DNMT3A y

DNMT3B junto con el factor DNMT3L, que recluta a las metiltransferasas al DNA mediante su interacción con la histona H3.

Durante la reprogramación epigenética que tiene lugar en la gametogénesis no sólo se producen cambios en los patrones de metilación, sino que la estructura de la cromatina también sufre importantes remodelaciones, sobre todo a nivel de la histona H3, entre las que se produce la eliminación de marcas de heterocromatina como la trimetilación de las lisinas 9 y 27 (H3K9me3 y H3K27me3) o el incremento en los niveles de marcas de cromatina activa como la di- y trimetilación de la lisina 4 (H3K4me2/3) o la acetilación de la lisina 9 (H3K9ac) (Hajkova *et al.*, 2008).

La segunda reprogramación epigenética del genoma tiene lugar en las primeras etapas del desarrollo embrionario. En el momento de la fertilización se combinan en una única célula dos genomas epigenéticamente muy diferentes que deben reprogramarse para permitir la pluripotencia y el desarrollo del embrión. Esta reprogramación epigenética se inicia con una desmetilación global, la cual comienza justo después de la fertilización y alcanza los niveles mínimos de metilación en la fase de blastocisto. Sin embargo, la velocidad a la que se desmetilan los genomas materno y paterno es diferente. El genoma materno parece desmetilarse por un mecanismo pasivo, es decir, la metilación se va perdiendo a medida que se suceden las divisiones celulares debido a que la replicación del DNA en ausencia de una actividad metiltransferasa que mantenga los patrones de metilación. El genoma paterno, en cambio, se desmetila prácticamente por completo en la primera división celular, lo que indica que debe existir un mecanismo de desmetilación activa (Weaver *et al.*, 2009) que, al igual que en la desmetilación que tiene lugar en los PGCs, parece estar ligada a la aparición de roturas de cadena simple y la actividad de la maquinaria de BER (Hajkova *et al.*, 2010). Sin embargo, a diferencia de los que ocurre en los PGCs, las marcas de *imprinting* establecidas durante la gametogénesis sobreviven a este nuevo proceso de desmetilación global. El mecanismo exacto por el que estas regiones escapan a la desmetilación global que sufre el genoma no se conoce en detalle, pero se ha propuesto la implicación de las metiltransferasas DNMT1 y DNMT1o, así como de otros factores proteicos como son MBD3, STELLA o ZFP57 (Weaver *et al.*, 2009).

Tras la desmetilación global del genoma se produce una oleada de metilación *de novo* que es llevada a cabo por las DNA metiltransferasas *de novo*, DNMT3A y DNMT3B. El patrón de metilación establecido en este momento será mantenido a lo largo de las divisiones celulares subsiguientes durante la vida del individuo por la DNA metiltransferasa de mantenimiento DNMT1 en combinación con UHRF1, aunque numerosas evidencias sugieren que DNMT3A y

DNMT3B podrían contribuir al mantenimiento de los patrones de metilación en determinadas regiones con elevada densidad de citosinas metiladas, como las regiones repetitivas (Jones & Liang, 2009).

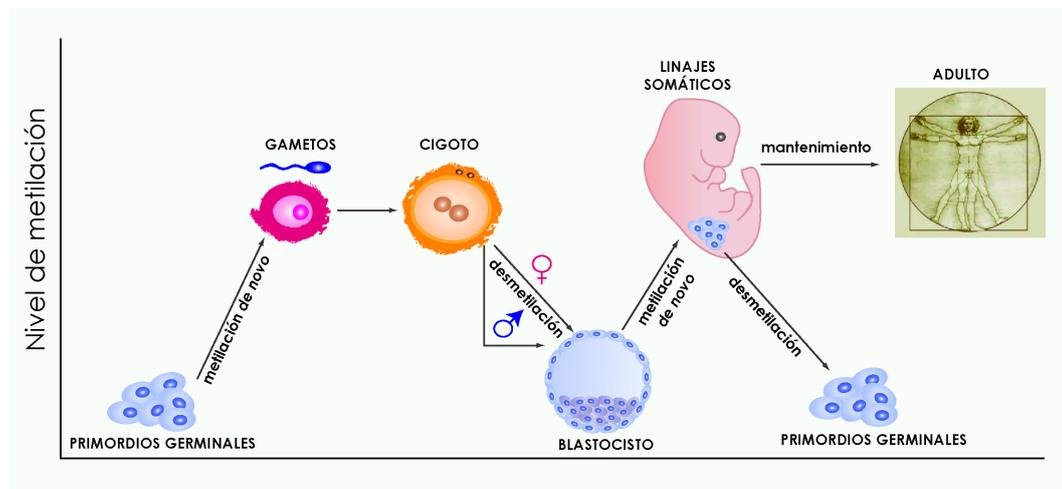


Figura 2. Establecimiento y mantenimiento de los patrones de metilación en el genoma de mamíferos

1.5. Funciones de la metilación genómica en mamíferos

La metilación del DNA se asocia generalmente a la represión transcripcional, aunque son numerosas las funciones biológicas que se le atribuyen a esta modificación del DNA. La metilación de la región promotora de un gen provoca el silenciamiento transcripcional del mismo y existen varios mecanismos por los que la metilación del DNA puede provocar esta represión. En primer lugar, hay evidencias de que al metilarse el DNA se impide la unión de determinados factores proteicos reguladores de la transcripción, como es el caso de CTCF, relacionado con las barreras entre dominios transcripcionales, que se une al DNA no metilado y es capaz de aislar a un promotor de la influencia de secuencias reguladoras remotas siendo de gran importancia en la regulación de la expresión de varios genes de *imprinting* (Filippova, 2007).

El segundo mecanismo de represión de la transcripción por la metilación del DNA está mediado fundamentalmente por proteínas de unión a DNA metilado que reclutan complejos remodeladores de la cromatina y represores transcripcionales (Ver apartado 1.3). No obstante, existen evidencias de que las propias DNMTs son capaces, no sólo de metilar el DNA sino, además, de reclutar complejos represores mediante su interacción con metiltransferasas y desacetilasas de histonas (Geiman *et al.*, 2004).

Sin embargo, y pese a su indiscutible implicación en la represión transcripcional, la metilación del DNA no parece ser el mecanismo principal de silenciamiento génico de la célula, sino que constituye un mecanismo de silenciamiento estable de regiones previamente silenciadas por otros medios. En el caso del cromosoma X inactivo en hembras de mamíferos la inactivación se inicia como consecuencia de la expresión de un RNA antisentido a partir del gen *XIST* y se estabiliza posteriormente mediante la metilación del DNA (Wutz & Jaenisch, 2000). Otro ejemplo es el silenciamiento de genes asociados al mantenimiento de la pluripotencia, como *OCT4*, o la represión de los satélites pericentroméricos. En ambos casos se ha demostrado que las modificaciones histónicas represoras son previas a la metilación del DNA y contribuyen al reclutamiento de las DNMTs (Feldman *et al.*, 2006). Otro dato que apunta a que la metilación del DNA no constituye el principal mecanismo de represión transcripcional de la célula es el hecho de que muchos promotores permanezcan desprovistos de metilación a pesar de que el gen no se esté expresando (Illingworth *et al.*, 2008; Rauch *et al.*, 2009)

Otra de las funciones atribuidas a la metilación genómica en mamíferos es la protección frente a los elementos transponibles (Yoder *et al.*, 1997), los cuales constituyen aproximadamente un 45% del genoma humano y se encuentran densamente metilados. La transcripción de estos elementos resultaría altamente perjudicial para el genoma, no sólo por la mutagénesis derivada de su movilidad en el genoma, sino además por la generación de un ruido transcripcional de niveles inaceptables que, seguramente, interferiría con el correcto funcionamiento del programa de expresión de la célula (Bird, 1995). De acuerdo con esto se ha visto que deficiencias en distintas DNMTs provocan la reactivación de retrotransposones en embriones (Walsh *et al.*, 1998) y línea germinal (Bourc'his *et al.*, 2001; Bourc'his & Bestor, 2004).

Por otra parte, numerosos estudios han puesto de manifiesto que la metilación del DNA es importante a la hora de mantener la estabilidad genómica. Defectos en las DNMTs conducen frecuentemente a cambios en la ploidía de las células y aberraciones cromosómicas (Dodge *et al.*, 2005; Karpf & Matsui, 2005) y se ha observado que la pérdida de metilación en la región de los telómeros coincide con un incremento en la frecuencia de recombinación (Karpf & Matsui, 2005).

Finalmente, se ha propuesto que la metilación del DNA podría contribuir al mantenimiento de la estructura de la cromatina después de la replicación del DNA (Cedar & Bergman, 2009). Evidencias recientes sugieren que el establecimiento de los patrones de metilación en fases tempranas del desarrollo podría estar mediado por modificaciones histónicas (Ooi *et al.*, 2007). Según este modelo el patrón de modificaciones histónicas debe formarse en el embrión en una fase previa a la metilación de *ново*, de modo que los patrones

de metilación se establecieron en base a la estructura de la cromatina. De este modo el patrón de metilación, que se mantiene a lo largo de las distintas divisiones celulares a través de la DNMT1, contribuiría a su vez a la herencia de las modificaciones de histonas. Así, mediante las proteínas de unión específica a DNA metilado se reclutarían complejos remodeladores de la cromatina con desacetilasas y metiltransferasas de histonas que establecerían las marcas apropiadas manteniendo a la cromatina en una conformación compacta y poco accesible.

1.6. Distribución de la metilación y organización de los dinucleótidos CpG en el genoma de mamíferos

El contenido en CpGs del genoma de mamíferos está en torno a un 20% de los CpGs que cabría esperar en base a su composición nucleotídica. Esta supresión en CpGs se debe a la elevada frecuencia de mutación de las metilcitosinas por desaminación espontánea (Holliday & Grigg, 1993) (**Fig. 3**). Esta desaminación

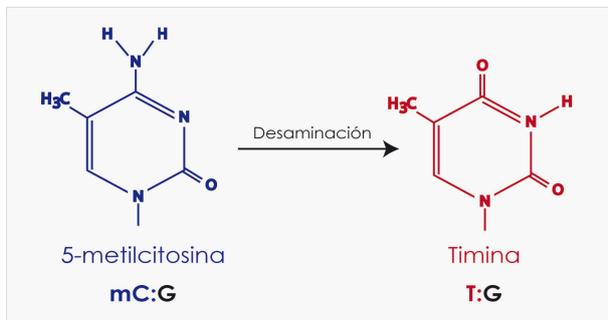


Figura 3. Desaminación espontánea de las citosinas metiladas

ocurre con una frecuencia de dos a cuatro veces superior a la de la citosina no metilada. Además la desaminación de la citosina da lugar a uracilo, una base extraña en el DNA, que es eficazmente eliminada por la uracilo DNA glicosilasa (UDG). Sin embargo, al desaminarse la metilcitosina se genera timina, una de las bases que componen normalmente el DNA y por tanto, a pesar de que también existen proteínas como TDG y MBD4 que reparan específicamente estos emparejamientos erróneos, su reparación es menos eficiente y con cierta frecuencia el emparejamiento T:G se corrige manteniendo la T de modo que el CpG pasa a ser TpG (Pfeifer, 2006). El hecho de que los CpGs no hayan desaparecido del genoma en su totalidad se debe a que se ha alcanzado un equilibrio entre la tasa de pérdida de CpGs por desaminación espontánea y su generación *de novo* por mutación puntual a partir de otros dinucleótidos (Sved & Bird, 1990).

Entre un 70 y un 80% de los dinucleótidos CpGs se encuentran metilados en mamíferos (Lister *et al.*, 2009), no obstante la distribución de CpGs metilados y no metilados no se produce de manera homogénea a lo largo del genoma. La mayor parte de los CpGs no metilados se encuentran agrupados en regiones que tienen un tamaño promedio de 1 Kb y reciben el nombre de islas CpG debido a su elevado contenido en este dinucleótido con respecto al resto del genoma.

2. ISLAS CpG

Las islas CpG constituyen la fracción no metilada del genoma. Fueron identificadas por primera vez mediante la digestión del DNA genómico de varios vertebrados con la enzima *HpaII*, que reconoce el sitio CCGG y es sensible a la metilación. Se observó que se generaba una población de fragmentos pequeños de DNA que procedían de secuencias que contenían *clusters* de CpGs no metilados y que en su conjunto constituían aproximadamente un 1% del genoma de mamíferos (Cooper *et al.*, 1983; Bird *et al.*, 1985). Su riqueza en dinucleótidos CpG con respecto al resto del genoma hace que sean fácilmente identificables representando los CpGs sobre la secuencia del DNA.

2.1. Propiedades de las islas CpG

La elevada densidad de CpGs que caracteriza a las islas CpG se debe a que estas regiones poseen un elevado contenido en los nucleótidos guanina(G) y citosina(C), superior al 65% en humanos, frente al promedio genómico que se sitúa en torno al 40% (Antequera, 2003) (**Fig. 4A**). Sin embargo, existen en el genoma otras regiones ricas en G+C que no presentan una alta densidad de CpGs. La diferencia radica en que las islas están libres de metilación y, por tanto, escapan a la supresión de CpGs por desaminación espontánea a la que está sometido el resto del genoma. Por ello, las islas presentan la frecuencia de CpGs que cabría esperar en base a su composición nucleotídica, mientras que la fracción metilada presenta sólo un 20% de los CpGs esperados (**Fig. 4A**).

Se han generado múltiples algoritmos para anotar las islas CpG en el genoma (Zhao & Han, 2009). Inicialmente las islas se definieron como regiones de al menos 200 pares de bases de longitud, con un contenido mínimo en G+C del 50% y una frecuencia de CpGs (CpG O/E: relación entre CpGs observados y esperados) de al menos 0,6 (Gardiner-Garden & Frommer, 1987). Posteriormente, se observó que aumentando el tamaño mínimo a 500 pares de bases se excluían la mayor parte de las repeticiones del tipo Alu que, por su alto contenido en G+C, eran frecuentemente identificadas como islas por el algoritmo anterior (Takai & Jones, 2002). Actualmente estos algoritmos siguen siendo los más utilizados, en combinación con bases de datos de elementos repetitivos, para identificar las islas CpG.

Las islas CpG difieren del resto del genoma no sólo en la composición nucleotídica sino también en la organización de la cromatina. Presentan altos niveles de acetilación de las histonas H3 y H4, así como una gran reducción en los niveles de la histona H1 (Tazi & Bird, 1990). Por otra parte, mediante análisis de ChIP-on-Chip con anticuerpos específicos se ha observado

que en las islas la histona H3 presenta altos niveles de metilación de la lisina 4 (Barski *et al.*, 2007; Weber *et al.*, 2007) y de acetilación de las lisinas 9 y 14 (Roh *et al.*, 2005), ambas marcas de cromatina transcripcionalmente activa (Guenther *et al.*, 2007). Este tipo de análisis también ha permitido determinar que las islas carecen de marcas represoras como la metilación de la lisina 36 de la histona H3 (Blackledge *et al.*, 2010). En las islas CpG los nucleosomas se ensamblan de forma inestable (Ramirez-Carrozzi *et al.*, 2009) de modo que es frecuente encontrar regiones libres de nucleosomas (Tazi & Bird, 1990). Esta organización convierte a las islas en regiones muy accesibles del genoma, lo cual concuerda con el hecho de que las islas CpG colocalizan con el 70% de los promotores humanos (Larsen *et al.*, 1992) y funcionan frecuentemente como orígenes de replicación (Giacca *et al.*, 1994; Delgado *et al.*, 1998; Antequera, 2004) (**Fig. 4B**).

Una relación directa entre la composición de bases de las islas y la estructura de la cromatina se ha establecido con el reciente descubrimiento de dos proteínas, CFP1 (Thomson *et al.*, 2010) y KDM2A (Blackledge *et al.*, 2010), que se unen específicamente a las islas CpG (**Fig. 4B**). Estas proteínas contienen un dominio CXXC que se une de forma específica a CpGs no metilados (Lee *et al.*, 2001). Lo que resulta más interesante es que ambas proteínas están relacionadas, directa o indirectamente, con actividades modificadoras de histonas. CFP1 forma parte del complejo Set1, el cual lleva a cabo la trimetilación de la lisina 4 de la histona H3, mientras que KDM2A es una desmetilasa que elimina activamente la metilación de la lisina 36 de la histona H3.

2.2. Islas CpG y promotores

Como se menciona en el apartado anterior, más del 70% de los genes humanos presentan una isla CpG en su promotor. Entre estos genes se incluyen todos los genes de expresión constitutiva y aproximadamente la mitad de los genes con expresión específica de tejido (Larsen *et al.*, 1992; Antequera & Bird, 1993; Zhu *et al.*, 2008). Esta elevada colocalización unida a la estructura permisiva de la cromatina en las islas, indica que estas podrían tener un papel en la regulación transcripcional. Esta conexión entre islas y promotores contrasta con el hecho de que no todas las islas CpG colocalizan con promotores de genes anotados. Es más, aproximadamente el 50% de las islas CpG se sitúan en regiones intergénicas e intragénicas. Sin embargo, estudios recientes han demostrado que estas islas, conocidas como islas huérfanas, poseen en su mayoría características de promotores funcionales. Así, un amplio porcentaje de ellas recluta a la RNAPol II, está enriquecido en H3K4me3 o presenta sitios de inicio de la transcripción en algún tejido (Illingworth *et al.*, 2010; Maunakea *et al.*, 2010; Medvedeva *et al.*, 2010). Estos datos sugieren que todas las islas CpG coinciden con promotores y que muchas de

las islas huérfanas son promotores de genes que no se habían caracterizado previamente, promotores alternativos de genes anotados o generan RNAs no codificantes.

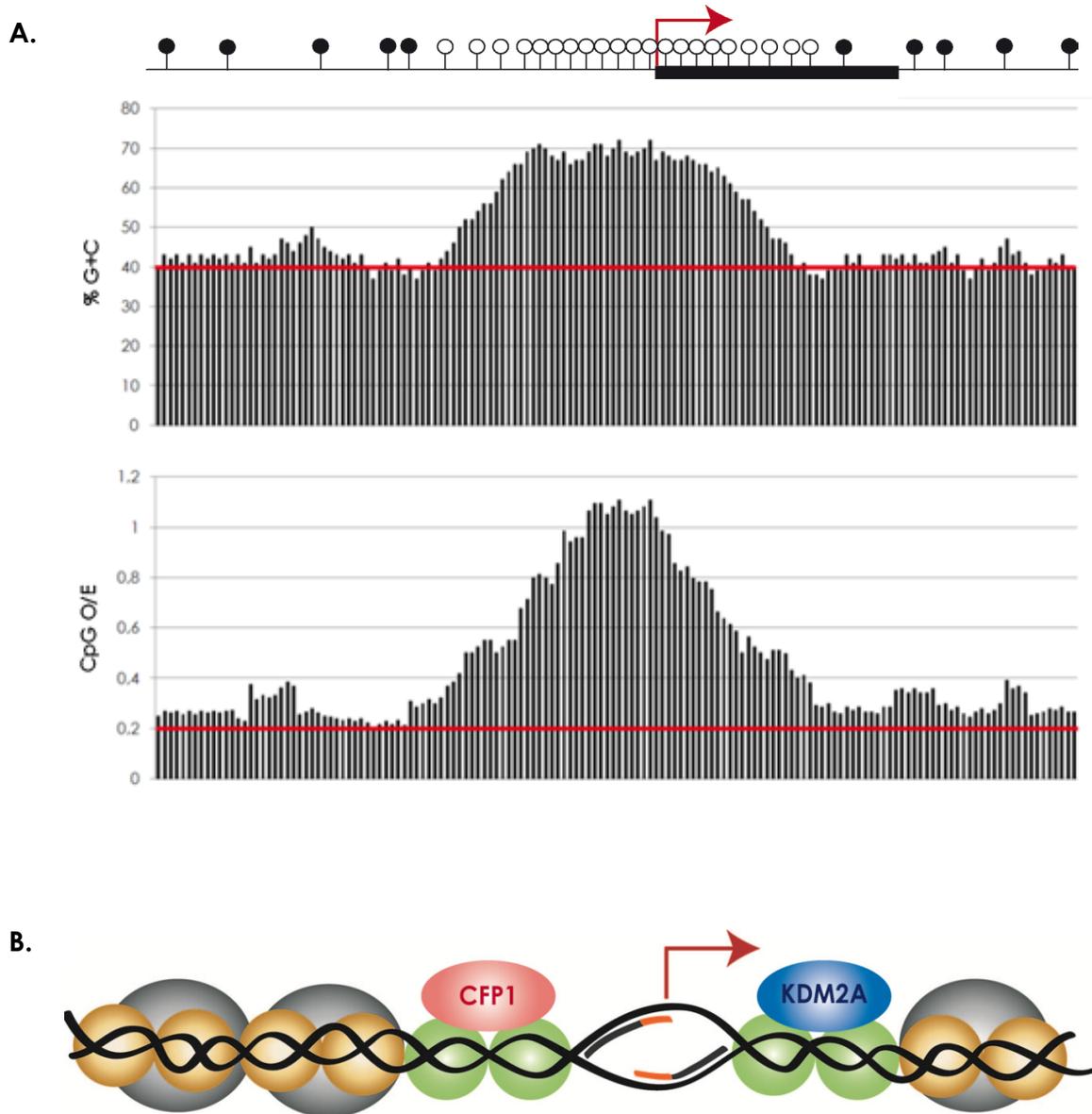


Figura 4. Propiedades de las islas CpG. A. Contenido en G+C y frecuencia de CpGs (CpG O/E: relación entre CpGs observados y esperados) en una isla modelo. La línea roja representa el valor promedio del genoma para estos parámetros. **B.** Estructura de la cromatina en las islas CpG. Se representan en verde los nucleosomas con modificaciones activas de histonas y en amarillo aquellos que presentan modificaciones represoras. Las esferas grises corresponden a proteínas de unión a DNA metilado (MBDs). CFP1 y KDM2A se unen de forma específica a los CpGs no metilados presentes en las islas CpG y contribuyen al mantenimiento de las modificaciones epigenéticas características de estas regiones.

Las diferencias entre los promotores con o sin isla CpG no se limitan a la secuencia. Se ha observado que los promotores asociados a islas CpG muestran múltiples sitios de inicio de la transcripción dispersos a lo largo de la isla mientras que los promotores que carecen de isla CpG suelen presentar un único sitio de inicio de la transcripción, generalmente asociado a cajas TATA (Carninci *et al.*, 2006). Los promotores asociados a islas CpG funcionan frecuentemente como promotores bidireccionales (Trinklein *et al.*, 2004), de hecho se ha descrito que en la mayor parte de los promotores asociados a islas CpG se produce transcripción divergente de manera constante que, en ausencia de señales que permitan la elongación, da lugar a transcritos de pequeño tamaño (Core *et al.*, 2008; Seila *et al.*, 2008). Se ha sugerido que esta generación constante de transcritos incompletos podría predisponer las islas a la transcripción de modo que la respuesta a señales regulatorias se produjera más rápido.

2.3. Islas CpG y orígenes de replicación

Los orígenes de replicación (ORIs) en mamíferos no se especifican mediante una secuencia consenso y se ha comprobado que el complejo de reconocimiento del origen (ORC) no muestra afinidad preferente por ningún tipo de secuencias (Vashee *et al.*, 2003). Esto sugiere que la especificación de los ORIs en mamíferos podría estar mediada por factores epigenéticos.

Hace más de diez años que se observó que las islas CpG aparecían sobrerrepresentadas en las cadenas nacientes de pequeño tamaño, al realizar análisis de intermediarios de replicación *in vivo* (Delgado *et al.*, 1998). Esto sugería que las islas CpG funcionan frecuentemente como orígenes de replicación. Estudios más recientes en los que se ha combinado el análisis de intermediarios de replicación con la hibridación de *microarrays* han confirmado que en torno al 50% de los orígenes de replicación colocalizan con islas CpG (Cadoret *et al.*, 2008). Además se ha observado que los orígenes que presentan una mayor eficiencia de activación son aquellos asociados a islas CpG (Sequeira-Mendes *et al.*, 2009).

A pesar de la ausencia de un requerimiento de secuencia para la especificación de los ORIs en mamíferos, se ha observado que sí existe una conexión entre el inicio de la replicación y la transcripción. Son muchas las observaciones que conectan ambos procesos. Así, se ha descrito que la transcripción es capaz de modular la actividad de los ORIs (Mesner & Hamlin, 2005), que los factores de transcripción son capaces de interactuar con ORC (Saitoh *et al.*, 2002; Minami *et al.*, 2006) y que los ORIs asociados a promotores activos muestran una activación temprana durante la fase S (Farkash-Amar *et al.*, 2008; Sequeira-Mendes *et al.*, 2009).

Sequeira-Mendes *et al.* observaron, además, que los sitios de inicio de la replicación coinciden con los sitios de inicio de la transcripción en los ORIs asociados a promotores, lo que sugiere una evolución conjunta de ambos procesos. Todas estas evidencias indican que la transcripción activa puede tener un papel en la especificación y activación de los orígenes de replicación en el genoma de mamíferos. Este papel podría ser activo, mediante el reclutamiento directo de la maquinaria de replicación o bien pasivo, mediante el remodelamiento de la cromatina en estas regiones de modo que estén más accesibles para la unión de ORC. Las islas CpG poseen ambas propiedades, es decir, funcionan como promotores y poseen una cromatina accesible, de modo que tanto si el reclutamiento de la maquinaria replicativa a los promotores activos se produce de forma activa como si lo hace de forma pasiva, no es extraño que las islas CpG colocalicen con orígenes de replicación con mayor frecuencia que otras regiones del genoma.

Resulta interesante señalar que, del mismo modo que en los promotores asociados a islas se generan transcritos de pequeño tamaño, se ha descrito que en los orígenes de replicación existe un fenómeno de re-replicación o replicación abortiva que genera una población de fragmentos pequeños de doble cadena que coinciden con el sitio de inicio de la transcripción (TSS) del gen adyacente (Gómez & Antequera, 2008). En conjunto, ambos fenómenos sugieren que las islas CpG se encuentran en un estado de actividad constante que les permite iniciar rápidamente la transcripción o la replicación en presencia de las señales adecuadas.

2.4. Islas CpG y metilación

La mayor parte de las islas CpG permanecen libres de metilación incluso en aquellos tejidos en los que el gen asociado no se está expresando (Bird, 2002). Sin embargo, existen islas normalmente metiladas en condiciones fisiológicas así como islas que se metilan de forma aberrante en determinadas circunstancias. Entre las islas que se encuentran normalmente metiladas en el genoma de mamíferos están aquellas que se localizan en el cromosoma X inactivo de las hembras o las que se asocian con el alelo inactivo de los genes de *imprinting* (Edwards & Ferguson-Smith, 2007; Reik, 2007). En ninguno de estos casos la metilación de las islas CpG constituye el mecanismo primario de inactivación, sino que tiene lugar cuando la transcripción ya ha sido reprimida por otros medios. El reciente desarrollo de técnicas de análisis de metilación a nivel genómico (Laird, 2010) ha permitido determinar que existe además un porcentaje de islas CpG, no asociadas al cromosoma X ni a genes de *imprinting*, que se encuentran metiladas normalmente en tejidos somáticos (Yamada *et al.*, 2004; Weber *et al.*, 2005; Eckhardt *et al.*, 2006; Weber *et al.*, 2007; Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Rauch *et al.*, 2009; Straussman *et al.*, 2009; Illingworth *et al.*, 2010; Maunakea *et al.*, 2010). La proporción

de islas metiladas varía ampliamente (entre un 9 y un 25%) en función del estudio que consideremos. Por otra parte, estos análisis de metilación global han mostrado que los patrones de metilación son específicos de tejido. La mayor parte de los análisis globales de metilación coinciden en que las islas que se metilan con mayor frecuencia son aquellas que no se asocian a ningún promotor anotado, mientras que la mayor parte de islas CpG asociadas a promotores se mantienen libres de metilación.

Además de las islas CpG metiladas en condiciones normales, existen islas que sufren metilación aberrante en células somáticas provocando la inactivación del gen asociado. Es frecuente encontrar metilación aberrante asociada a procesos de envejecimiento así como al desarrollo de procesos tumorales. Durante el envejecimiento, se acumulan en las células alteraciones epigenéticas que se observan como una disminución en los niveles globales de metilación y un aumento en la metilación aberrante de islas CpG (Fraga & Esteller, 2007; Maegawa *et al.*, 2010). Esta desregulación epigenética ha sido atribuida por algunos autores a la alteración de la expresión de las DNMTs, con una disminución de la DNMT1, que explicaría el descenso en los niveles de metilación globales y un incremento en la expresión de la DNMT3B, que justificaría una mayor frecuencia de metilación aberrante de islas CpG (Casillas *et al.*, 2003).

En las células cancerosas también es frecuente encontrar una reducción de la metilación global, así como una hipermetilación de islas CpG. Se han descrito numerosos casos en los que el desarrollo tumoral se asocia a la metilación aberrante del promotor de genes supresores tumorales, como Rb, p16^{INK4a} o RASSF1A, entre muchos otros (Esteller, 2007). La acumulación de alteraciones epigenéticas a lo largo del envejecimiento aumenta la probabilidad de que dichas alteraciones afecten a la expresión de genes supresores tumorales, lo que explica que la incidencia del cáncer aumente con la edad (Fraga *et al.*, 2007).

En el genoma de mamíferos existen, por tanto, distintas categorías de islas metiladas. Así, encontramos islas constitutivamente metiladas, islas que se metilan de forma específica de tejido e islas que son más propensas a sufrir metilación aberrante. Está por determinar el mecanismo que permite a la célula diferenciar estas islas entre ellas así como de la mayoría de islas del genoma que se encuentran constitutivamente no metiladas. Tampoco se han esclarecido los factores implicados en la protección de las islas frente a la metilación.

Con respecto a la protección de las islas frente a la metilación se han propuesto distintas hipótesis (Antequera & Bird, 1999; Illingworth & Bird, 2009). Una explicación sencilla sería que la islas no constituyeran un sustrato accesible para las DNMTs, pero el hecho de que existan islas

que se metilen en condiciones normales, hace que esta hipótesis sea insostenible (Fig. 5A). Otra posibilidad sería que existiese un mecanismo activo de desmetilación que eliminase esta marca de las islas constantemente. Sin embargo, hasta la actualidad no se ha descrito ninguna actividad desmetilasa en tejidos somáticos de mamíferos (Fig. 5B).

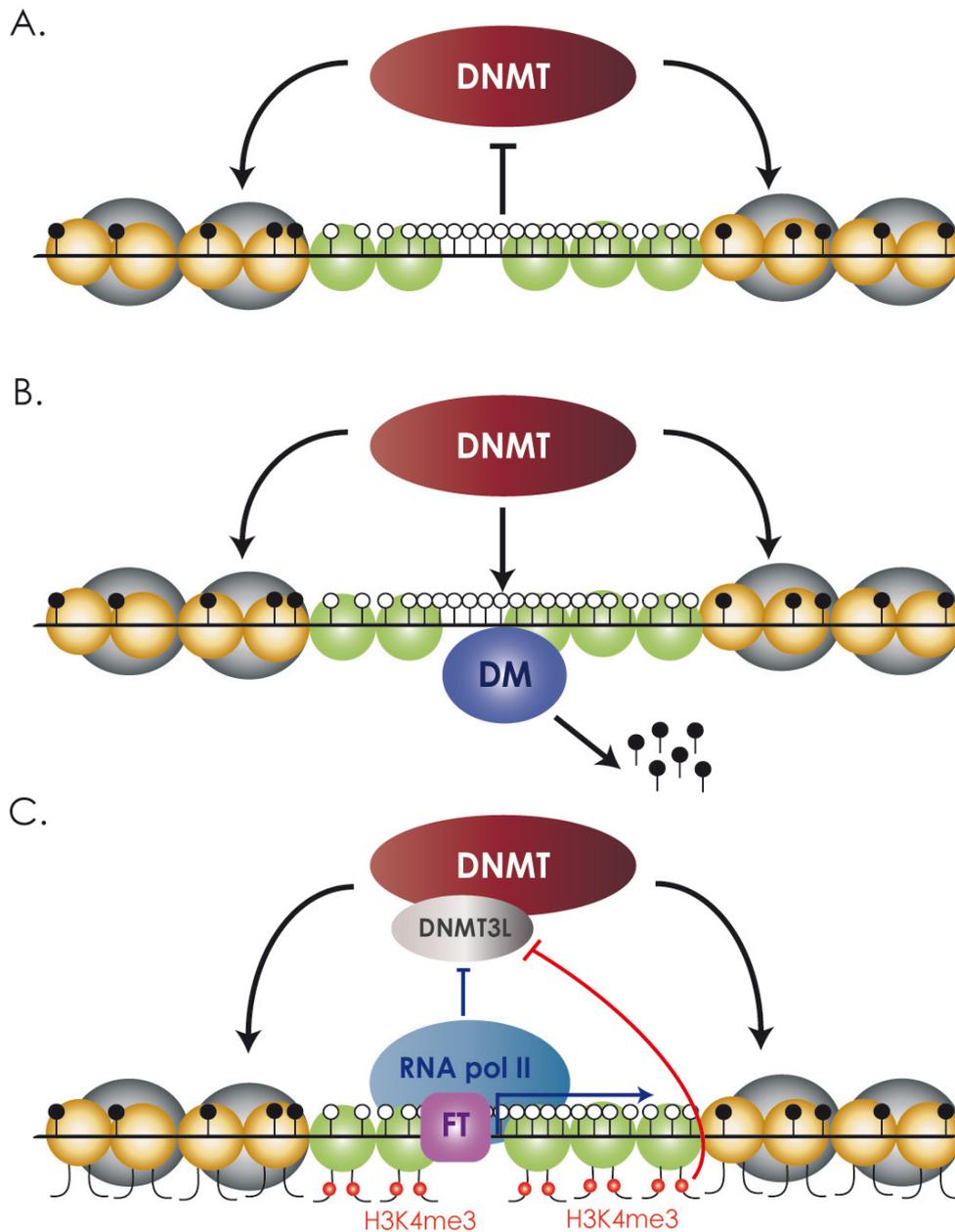
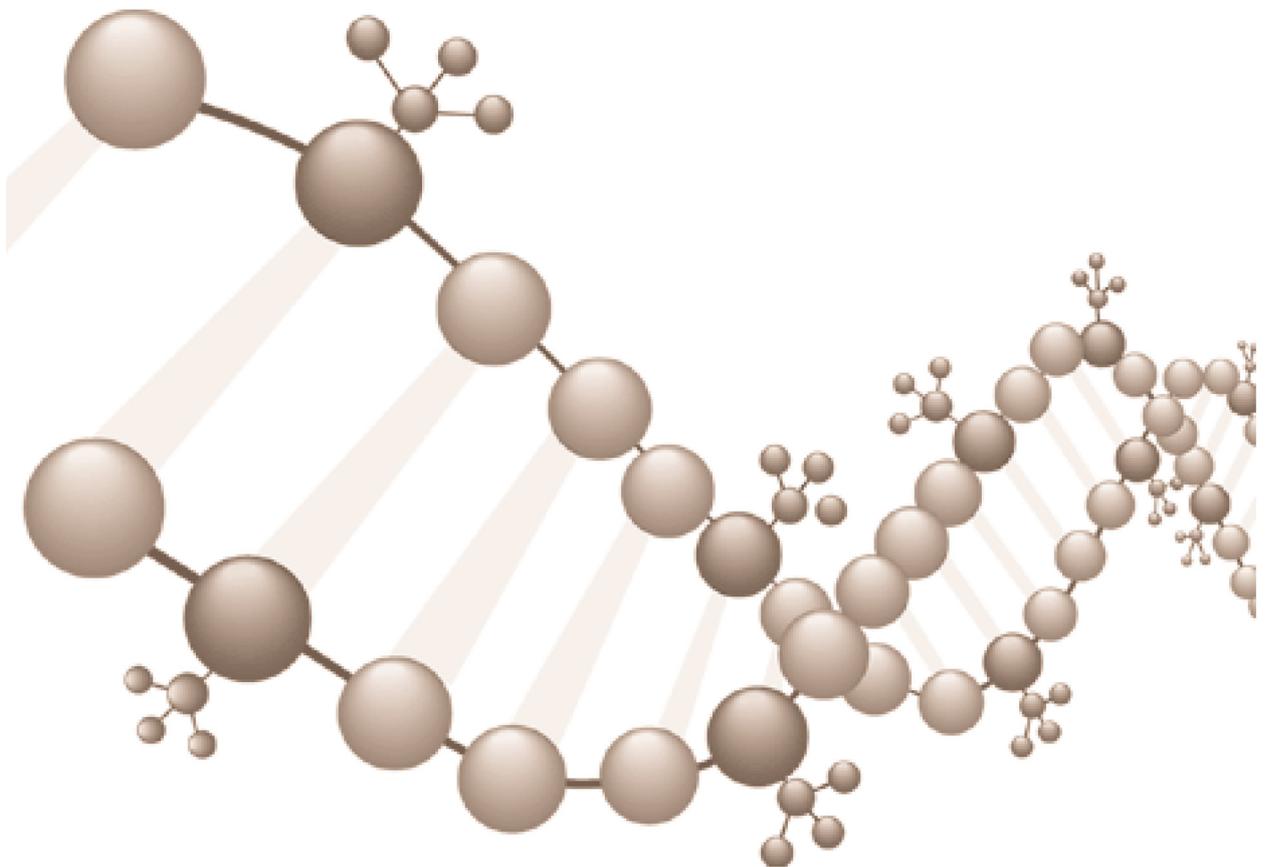


Figura 5. Distintas hipótesis sobre la protección de las islas CpG frente a la metilación. A. Inhibición directa de la unión por la secuencia de DNA. **B.** Desmetilación activa. **C.** Impedimento de la unión por la estructura de la cromatina, las modificaciones epigenéticas y las proteínas que interaccionan con las islas CpG.

Una tercera hipótesis, la más aceptada en la actualidad, es que la unión de las DNMTs se vea inhibida de forma directa o indirecta por la estructura de la cromatina y la unión de los factores de transcripción en estas regiones (**Fig. 5C**). De acuerdo con esta hipótesis se ha observado que la mutación de los sitios de unión de factores de transcripción de un promotor conlleva la metilación de la isla CpG (Macleod *et al.*, 1994). En línea con esto, la combinación de los análisis globales de metilación con los estudios a nivel genómico de localización de la RNAPol II, ha permitido determinar que la presencia de RNAPol II en una isla CpG es indicativa de la ausencia de metilación independientemente del nivel de expresión del gen asociado (Takeshima *et al.*, 2009). Por otra parte, los factores de transcripción permanecen unidos a las islas CpG asociadas a genes específicos de tejido incluso en los tejidos en los que el gen no se expresa (Cuadrado *et al.*, 2001). Esto explicaría que muchas islas CpG permanezcan libres de metilación en ausencia de la transcripción activa del gen asociado. Otro mecanismo de inhibición directa de la unión de las DNMTs a las islas CpG vendría de la mano de las modificaciones histónicas características de estas regiones. Como se ha descrito anteriormente, la mayor parte de las islas CpG no metiladas presentan altos niveles de H3K4me3. Estudios *in vitro* han demostrado que la unión del factor DNMT3L a la cromatina se produce a través de la cola de la histona H3, pero esta unión se inhibe en presencia de mono-, di- o trimetilación de la lisina 4 de esta histona (Ooi *et al.*, 2007). De acuerdo con esta hipótesis, las islas CpG no metiladas unen específicamente CFP1, que forma parte del complejo Set1, encargado de la trimetilación de la lisina 4 de la histona H3 (Thomson *et al.*, 2010).

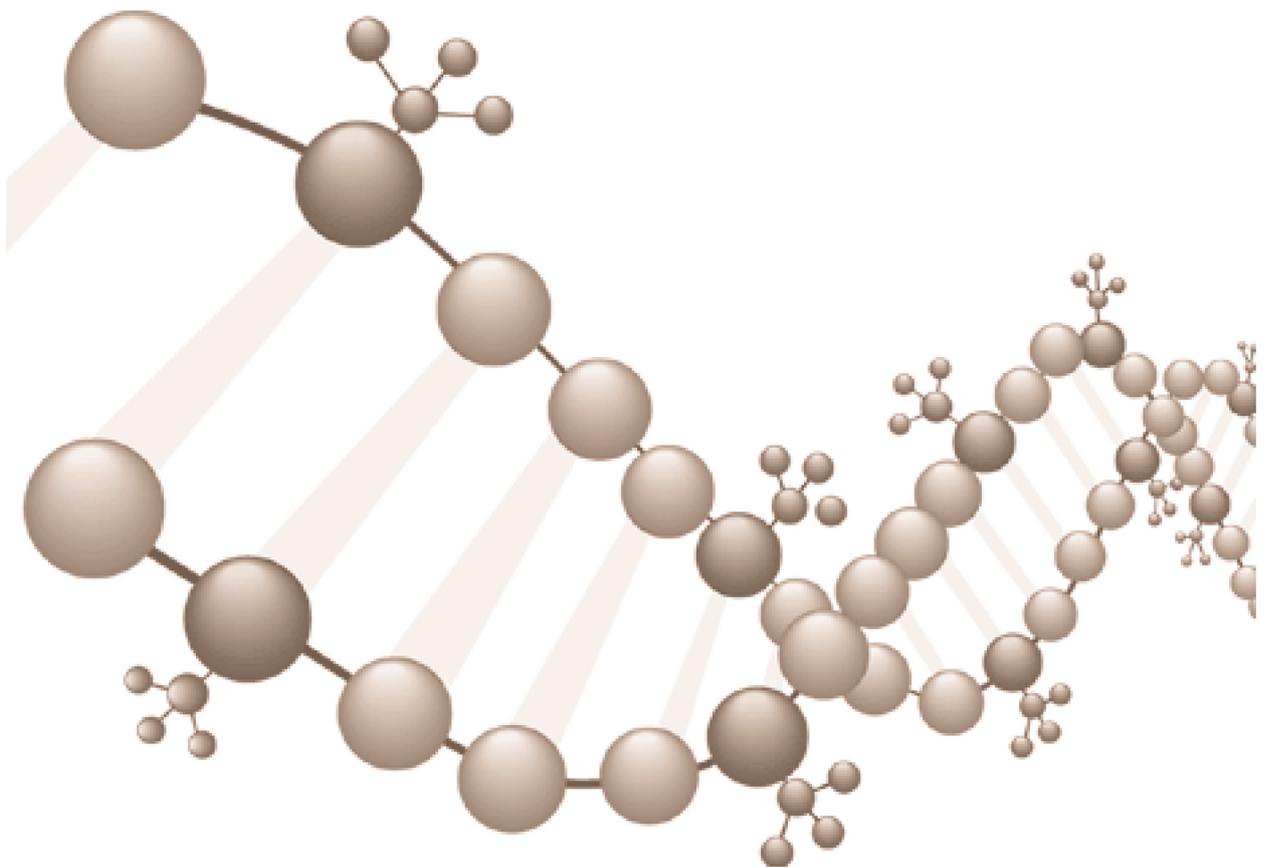
OBJETIVOS



El objetivo principal de esta tesis doctoral es contribuir a esclarecer cómo se han generado las islas CpG en el genoma de mamíferos y qué mecanismos permiten a estas regiones permanecer desprovistas de metilación. Para ello decidimos llevar a cabo los siguientes análisis:

1. Papel de la replicación del DNA en la protección frente a la metilación de las islas CpG.
2. Contribución de la transcripción al mantenimiento de las islas CpG en estado no metilado.
3. Dinámica de generación de las islas CpG en el genoma de mamíferos.

RESULTADOS



1. PAPEL DE LA REPLICACIÓN DEL DNA EN LA PROTECCIÓN FRENTE A LA METILACIÓN DE LAS ISLAS CpG

1.1 Antecedentes

Cuando se inició este trabajo existían en la literatura varias evidencias que sugerían que las islas CpG funcionan frecuentemente como orígenes de replicación en mamíferos. En nuestro laboratorio se demostró que las islas CpG se replican sincrónicamente al principio de la fase S en células CHO de hámster (Delgado *et al.*, 1998). Además, se observó que tres islas CpG seleccionadas al azar estaban presentes en una población de cadenas nacientes de DNA de pequeño tamaño, mientras que sus regiones flanqueantes estaban ausentes en dicha población. Por otra parte, en trabajos de otros grupos se observó que, al realizar experimentos de inmunoprecipitación de cromatina con anticuerpos específicos contra proteínas implicadas en el inicio de la replicación, más del 50% de los fragmentos inmunoprecipitados mostraban propiedades características de islas CpG (Keller *et al.*, 2002; Ladenburger *et al.*, 2002).

El número de ORIs que se habían descrito en mamíferos en el momento de iniciar esta tesis doctoral era escaso, sin embargo ya se apreciaba una tendencia a la colocalización entre ORIs e islas CpG (Giacca *et al.*, 1994; Keller *et al.*, 2002; Ladenburger *et al.*, 2002). Además parecía existir una tendencia de los ORIs que colocalizaban con islas a situarse en torno al extremo 5' de las mismas. Por otra parte, varias publicaciones (Ghazi *et al.*, 1992; Rideout *et al.*, 1994), así como observaciones realizadas en nuestro laboratorio (Cuadrado, Delgado & Antequera, sin publicar), sugerían que las islas CpG son más proclives a metilarse por el extremo 3' en líneas celulares y tumores, mientras que el extremo 5' de las mismas parecía más resistente a la metilación. Esta mayor protección frente a la metilación junto con la localización preferente de los ORIs en el extremo 5' de las islas nos llevó a barajar la hipótesis de que bien la actividad de inicio de la replicación o la maquinaria asociada a esta podrían estar participando en la protección de las islas frente a la metilación.

Otra observación interesante se deriva de un trabajo realizado en colaboración con el laboratorio del Dr. Manuel Serrano (CNIO, Madrid) en el que se observó que en la región promotora del gen *CDKN2B* humano existía un ORI, que denominaremos ORI *CDKN2B*, el cual mostraba un alto grado de conservación filogenética a pesar de tratarse de una región no codificante (Gonzalez *et al.*, 2006) (**Fig. 6**). El hecho de que la secuencia del ORI estuviera conservada sugería que podría existir también una conservación en la función de esta región entre los distintos organismos, lo que reforzaba la hipótesis de que el inicio de la replicación podía estar implicado en la protección frente a la metilación de las islas CpG en mamíferos.

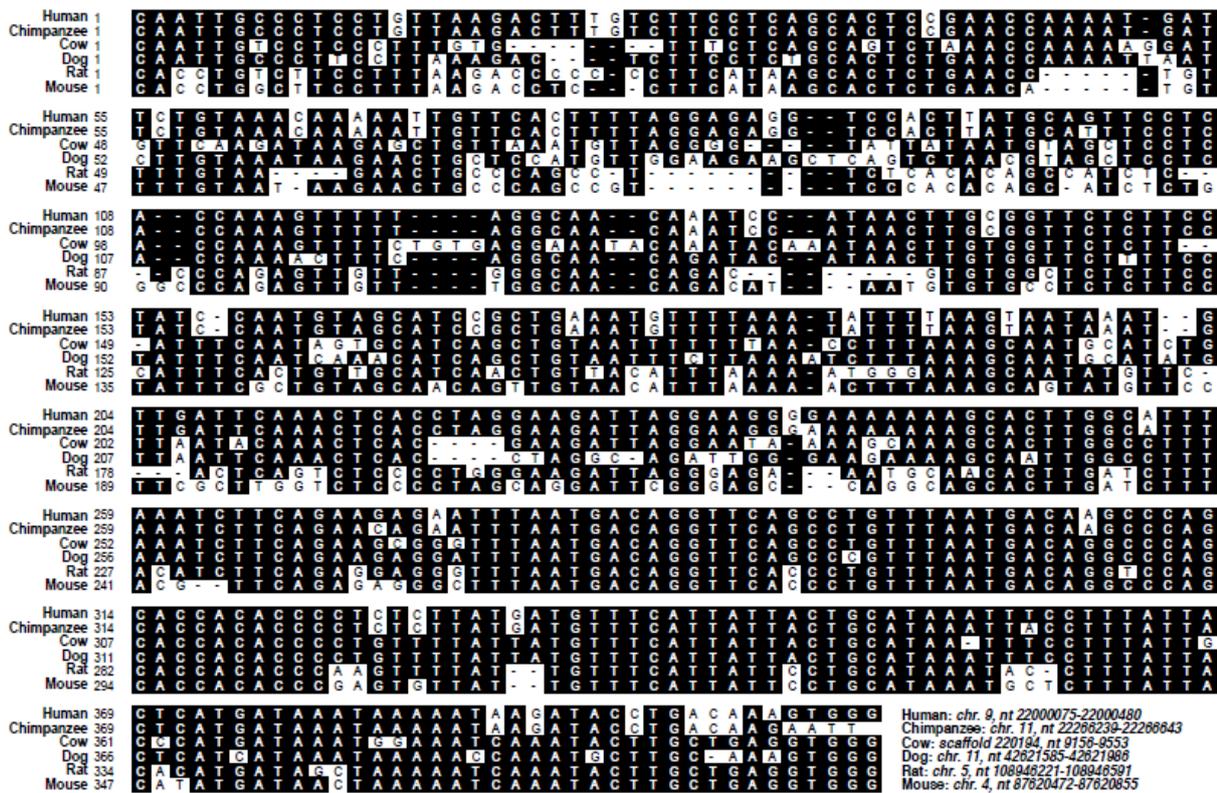


Figura 6. Alineamiento múltiple de la región del ORI *CDKN2B* en el genoma de varios mamíferos. Extraído de González et al., 2006.

1.2 Conservación filogenética de los ORIs de mamíferos que se encuentran asociados a islas CpG

En primer lugar nos propusimos determinar si el ORI *LMNB2*, uno de los ORIs mejor caracterizado en humanos, estaba filogenéticamente conservado, como se había observado en el ORI *CDKN2B*. Para ello se obtuvieron, a través del navegador UCSC Browser (Kent et al., 2002), las secuencias ortólogas de varios mamíferos (*Pan troglodytes*, *Macaca mulatta*, *Bos taurus*, *Mus musculus* y *Rattus norvegicus*) y se realizó un alineamiento múltiple utilizando el programa *Clustal X* (Chenna et al., 2003) y el editor de alineamientos *Jalview* (Waterhouse et al., 2009) (Fig. 7). Como se observa en la figura 7, la región del ORI mostraba un grado de conservación similar al del primer exón del gen adyacente a pesar de que el ORI *LMNB2* se encuentra en una región no codificante.

Tras comprobar que tanto el ORI *CDKN2B* como el ORI *LMNB2* se hallaban filogenéticamente conservados, nos planteamos si otros ORIs de mamíferos asociados a islas CpG también presentarían un nivel similar de conservación. Se seleccionaron varios de los ORIs

asociados a islas CpG que habían sido descritos hasta el momento (*TOP1*, *HSP70*, *C-FOS*, *GCLC*, *FUCA1*, *C-MYC*, *MCM4*) y se analizó su conservación filogenética mediante comparación entre secuencias ortólogas de varios mamíferos. Hemos denominado ORIs a aquellas regiones en las que se ha descrito el inicio de la replicación mediante el análisis de intermediarios de replicación en cadenas nacientes de DNA (Ver Materiales y Métodos). En el caso de los ORIs *TOP1*, *C-MYC* y *MCM4* se ha demostrado, además, la unión de alguna de las subunidades del complejo de reconocimiento del origen (ORC).

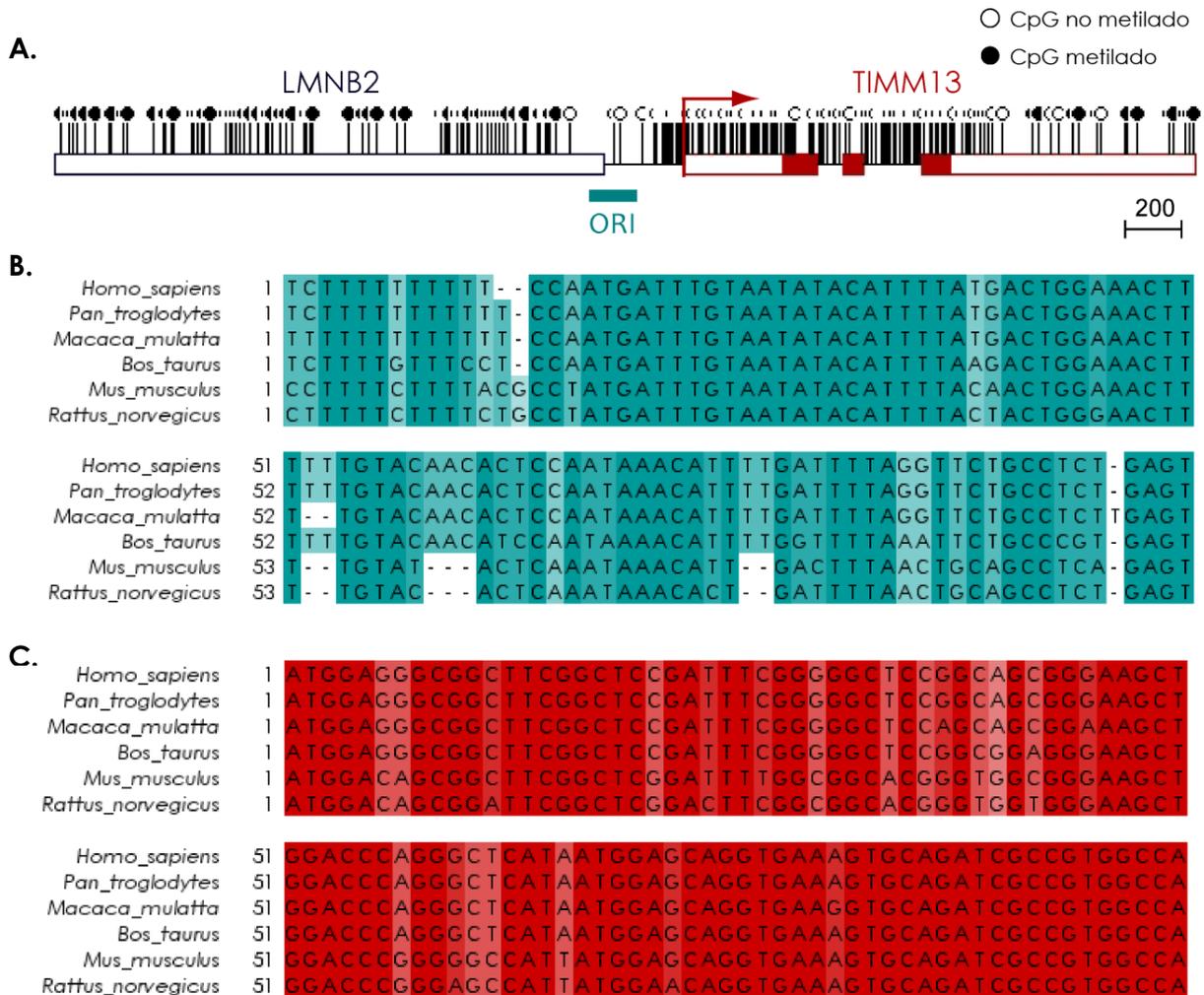


Figura 7. Alineamiento múltiple de la región del ORI LMNB2 en el genoma de varios mamíferos. A. Mapa de la región en la que se localiza el ORI LMNB2, entre el último exón del gen LMNB2 y el primer exón del gen TIMM13. Se representa mediante líneas verticales la posición de los dinucleótidos CpG, así como el estado de metilación de los mismos. **B.** Alineamiento múltiple de la región del ORI LMNB2 humano con su región ortóloga en los genomas de chimpancé (*Pan troglodytes*), macaco (*Macaca mulatta*), vaca (*Bos taurus*), ratón (*Mus musculus*) y rata (*Rattus norvegicus*) **C.** Alineamiento múltiple de la región codificante del primer exón del gen TIMM13 en las mismas especies.

Para analizar la conservación filogenética de estas regiones se utilizó, en primer lugar, el navegador *Vista Browser*, que nos permite visualizar el grado de conservación de una secuencia de referencia, en este caso la secuencia humana, con respecto a la región ortóloga en otras especies de mamíferos.

A continuación, y con el objetivo de confirmar la presencia o ausencia de elementos conservados en estas regiones, se utilizó el programa *Eshadow*, que además de realizar alineamientos múltiples estima la distancia filogenética que existe entre las secuencias analizadas, de modo que nos permite determinar qué regiones han divergido a una velocidad menor en un locus determinado, es decir, que regiones se encuentran significativamente conservadas (**Fig. 8 y Fig. S1 A-F**). Como puede observarse en la figura 8 y figura S1, seis de los nueve orígenes analizados, incluyendo *CDKN2B* y *LMNB2*, presentaban regiones con un alto grado de conservación filogenética en mamíferos a pesar de tratarse de regiones no codificantes. De este modo, y a pesar de que la muestra era pequeña, un 67% de los ORIs analizados asociados a islas CpG incluía regiones conservadas filogenéticamente en mamíferos.

Como se ha mencionado anteriormente, las islas CpG funcionan como orígenes de replicación. Por otra parte nuestros datos sugieren que los ORIs asociados a islas presentan frecuentemente conservación filogenética. Teniendo en cuenta ambas premisas, esperaríamos encontrar frecuentemente regiones conservadas no codificantes asociadas a islas CpG en mamíferos. Para intentar confirmar esta predicción analizamos 21Mb del genoma seleccionadas al azar y distribuidas en 7 bloques de 3 Mb cada uno (**Fig. 9B**). Este análisis se llevó a cabo utilizando dos navegadores, el *Vista Browser* (Frazer *et al.*, 2004) (ver **Fig. 8**) y el programa *PhastCons* (Siepel *et al.*, 2005) (**Fig. 9A**) integrado en el *UCSC Browser*. De este modo pudimos inspeccionar visualmente las regiones seleccionadas y determinar qué porcentaje de las islas CpG contenidas en ellas presentaban una región conservada en posición 5'. Sólo se incluyeron en el análisis las islas CpG asociadas a promotores anotados. Mediante este análisis pudimos determinar que más del 40% de las islas analizadas mostraban una región conservada no codificante en su extremo 5' (**Fig. 9C**).

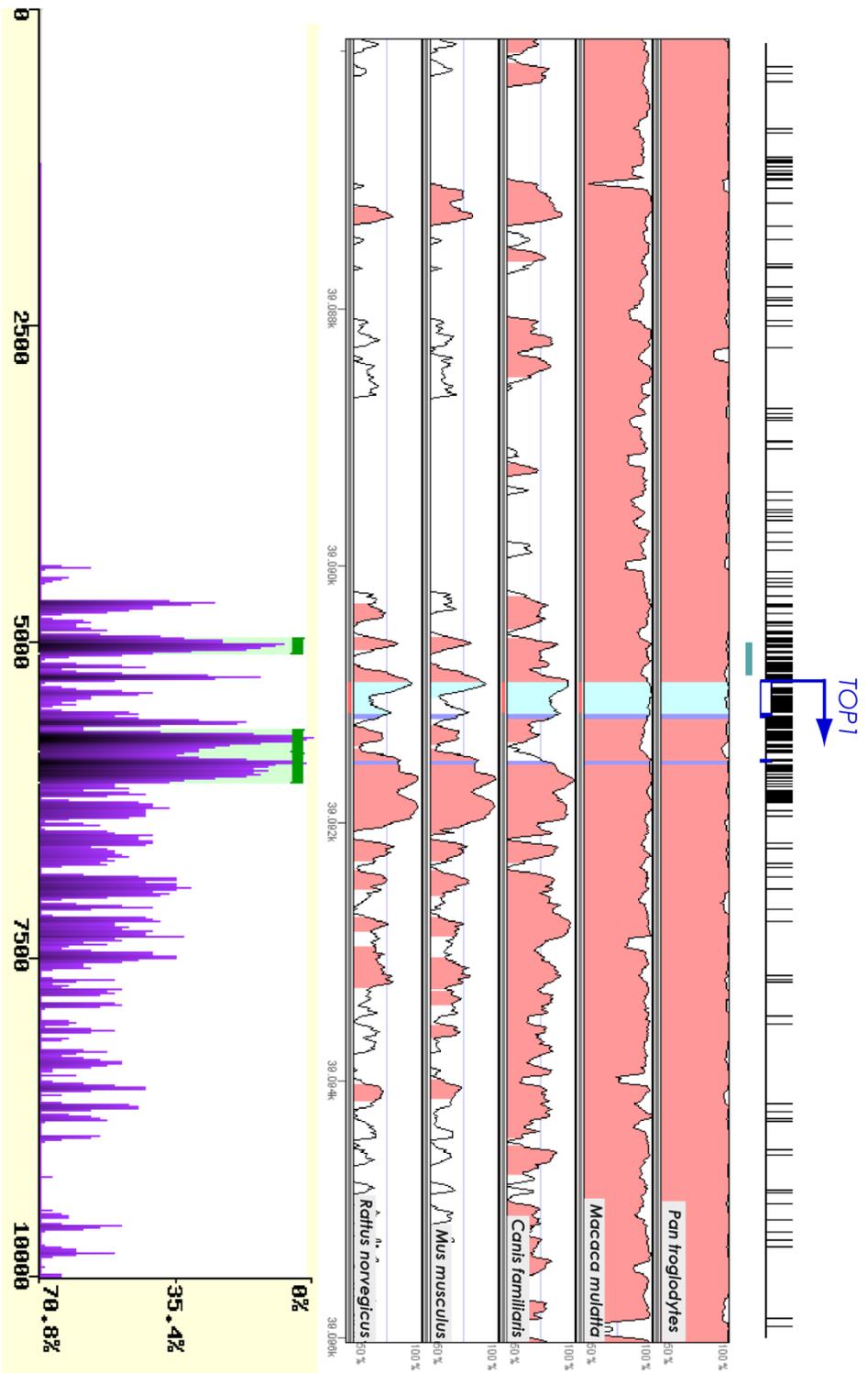


Figura 8. Análisis de conservación filogenética de ORIs asociados a islas CpG en el genoma de mamíferos. Análisis de la conservación filogenética del ORI asociado al gen *TOP1*. En la parte superior se encuentra el mapa de la región en el que se representa la región en la que se ha descrito el inicio de la replicación (línea verde), así como la posición de los dinucleótidos CpG (líneas verticales). En el centro se muestra el grado de conservación de la secuencia humana con respecto a la secuencia ortóloga de varios mamíferos obtenida a través del navegador *ViewBrowser*. Las regiones conservadas aparecen coloreadas en violeta cuando se trata de regiones codificantes y en rosado cuando son regiones no codificantes. En la parte inferior de la figura se muestra el grado global de divergencia en el conjunto de mamíferos analizados, estimado mediante el programa *Eshadow*. Se resaltan en verde aquellas regiones que, dado su bajo grado de divergencia, se consideran significativamente conservadas.

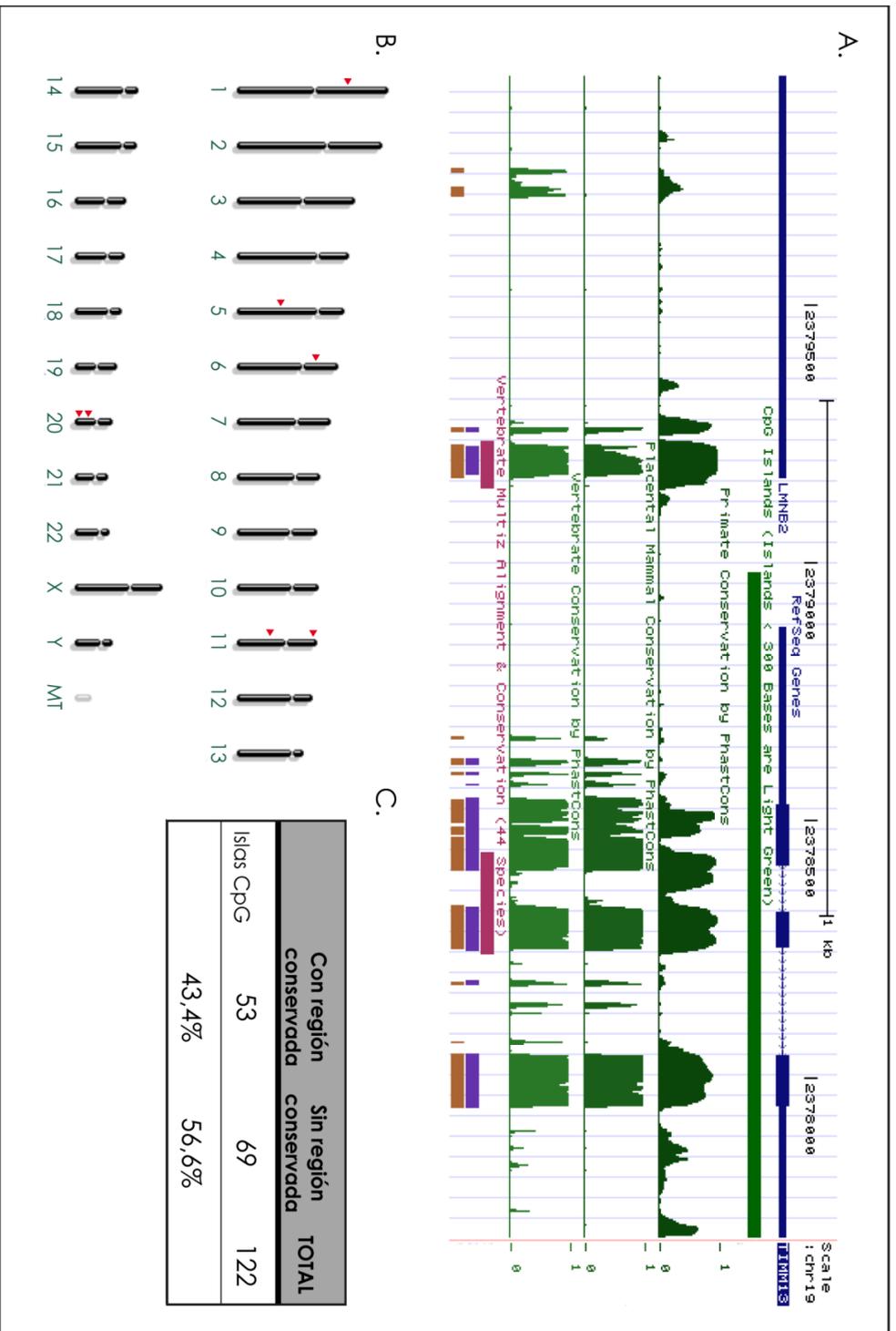


Figura 9. Rastreo bioinformático de regiones conservadas en el extremo 5' de islas CpG. **A.** Conservación del locus del ORI LMNB2 en primates, mamíferos y vertebrados obtenida mediante el programa PhastCons integrado en el UCSC Browser. **B.** Localización en el genoma humano de las siete regiones analizadas; de 3 Mb cada una (Chr1:65000000-68000000, Chr5: 114000000-117000000, Chr6:30000000-33000000, Chr11:1-3000000, Chr11:85000000-88000000, Chr20:43000000-46000000, Chr20:55000000-58000000). **C.** Recuento de islas CpG presentes en las siete regiones analizadas que poseen o no una región conservada en su extremo 5'.

1.3 Análisis de metilación de los ORIs *LMNB2* y *CDKN2B*

El análisis de metilación de la región que contiene el ORI *LMNB2* humano, llevado a cabo previamente al inicio de este trabajo (Cuadrado & Antequera, sin publicar), mostró que la región en la que se inicia la replicación establecía el límite de metilación en 5' de la isla CpG adyacente, asociada al promotor del gen *TIMM13* (Fig. 10).

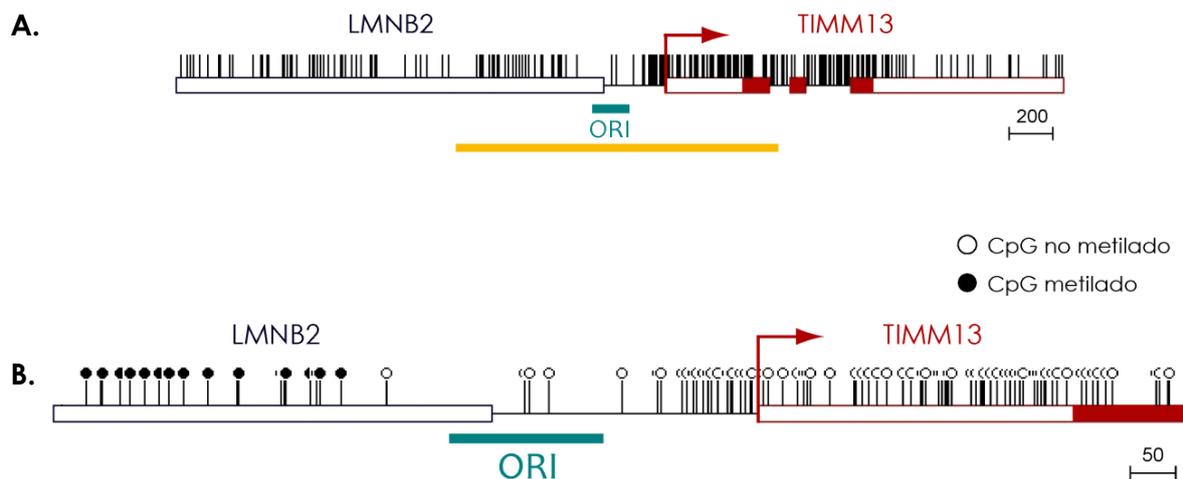


Figura 10. Análisis de metilación del ORI *LMNB2*. A. Mapa de la región en la que se encuentra el ORI asociado al último exón no codificante del gen *LMNB2*. Se indica con una línea amarilla la región cuyo patrón de metilación ha sido analizado. B. Mapa de metilación de esta región.

Con el objetivo de comprobar si el ORI *CDKN2B* establece el límite de metilación de la isla CpG adyacente en su extremo 5', como ocurre en el caso del ORI *LMNB2*, se llevó a cabo un análisis de metilación en DNA procedente de células HeLa y 293. Para ello se utilizó la técnica del bisulfito sódico, que permite conocer el estado de metilación de cada dinucleótido CpG en una región determinada del genoma. Se analizó la región comprendida entre los nucleótidos -1800 y +100 mediante PCR con oligonucleótidos específicos sobre el DNA transformado con bisulfito sódico. A continuación, se clonaron los productos de PCR en *E.coli* y se secuenciaron 20 y 10 clones del DNA de 293 y HeLa, respectivamente (Fig. 11B). El análisis de las secuencias obtenidas determinó que el límite de metilación de la isla en su extremo 5' coincidía con la región caracterizada como ORI, como se había observado en el caso del ORI *LMNB2* (Fig. 11C).

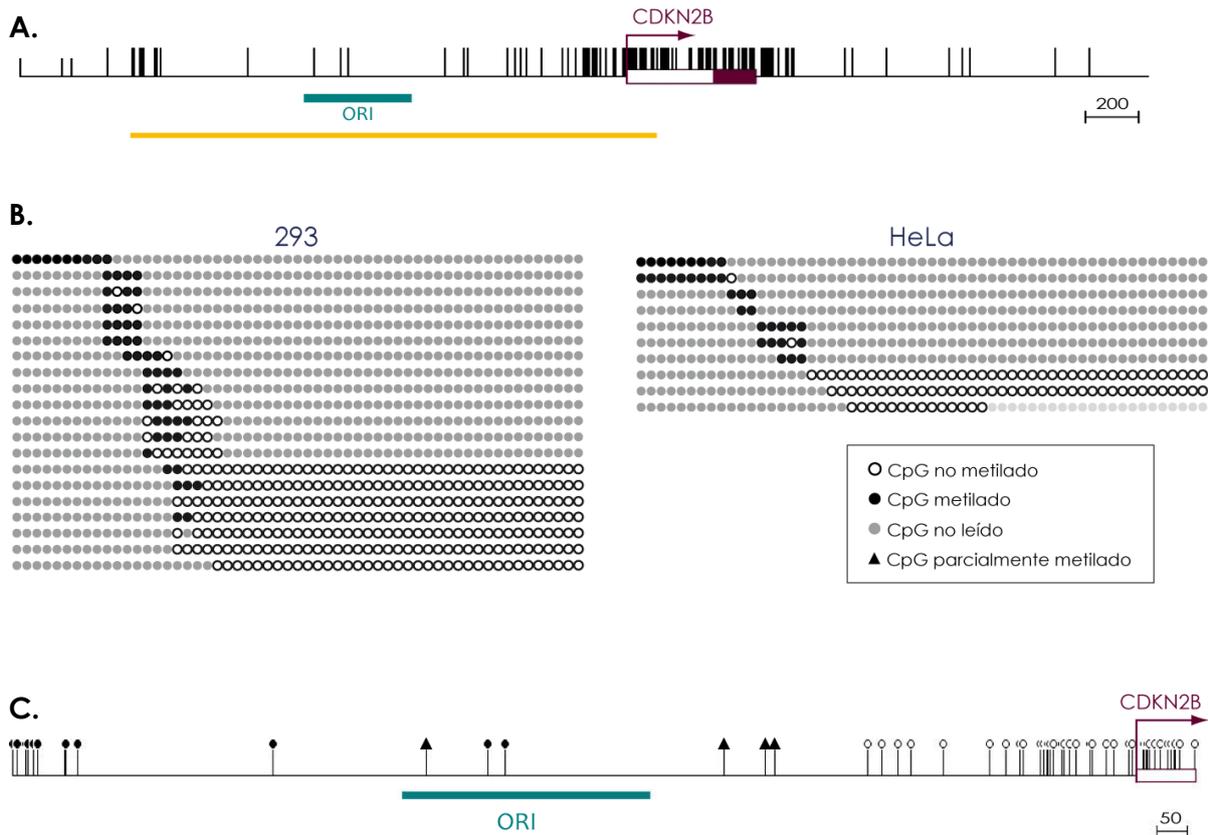


Figura 11. Análisis de metilación del ORI *CDKN2B*. **A.** Mapa del locus en el que se encuentra el ORI *CDKN2B*. Se indica con una línea amarilla la región cuyo patrón de metilación ha sido analizado. **B.** Patrón de metilación de los distintos clones analizados para cada línea celular. **C.** Patrón de metilación consenso obtenido tras promediar los patrones de los distintos clones secuenciados para la línea celular 293. Se especifica en el recuadro superior el significado de cada uno de los símbolos utilizados.

1.4 La unión de ORC como barrera frente a la metilación en el extremo 5' de las islas CpG

Teniendo en cuenta los datos obtenidos en los apartados anteriores, por los que las regiones asociadas a islas CpG en las que se inicia la replicación incluían frecuentemente regiones conservadas y un alto porcentaje de las islas CpG presentaba una región conservada en su extremo 5', nos preguntamos si podríamos detectar ORIs en base a la presencia de una región conservada no codificante en el extremo de una isla CpG y si la región conservada marcaría el límite de metilación. Para abordar esta cuestión se seleccionaron, al azar, cinco regiones conservadas no codificantes en el extremo 5' de islas CpG entre las detectadas en el análisis anterior (**Fig. 12 y Fig. S2 A-D**). El objetivo era determinar si estas regiones unían ORC y, de ser así, si establecían el límite de metilación de la isla asociada en su extremo 5', como sucedía en los dos orígenes de referencia de este estudio, *CDKN2B* y *LMNB2*.

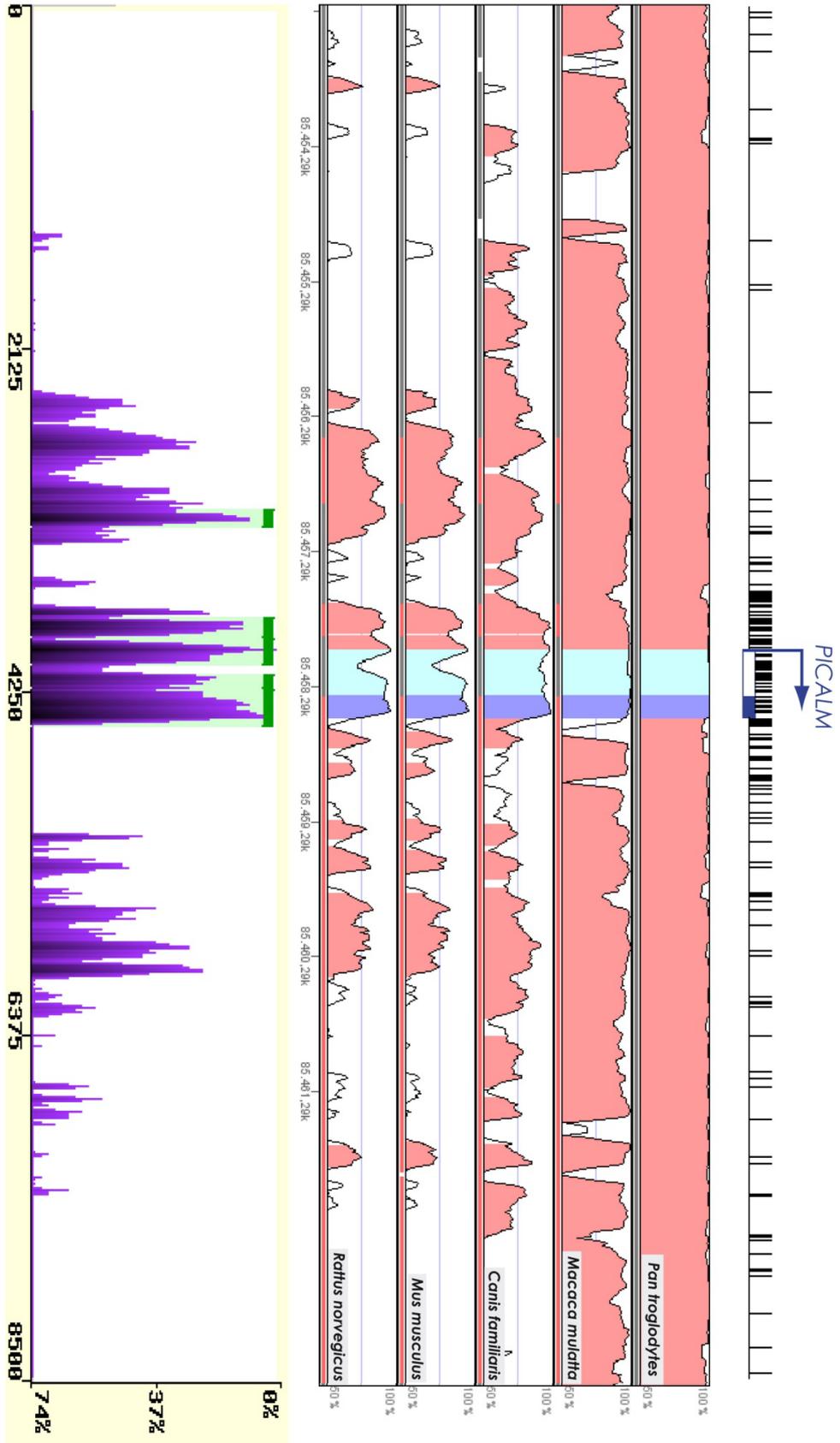


Figura 12. Análisis de conservación filogenética en el extremo 5' de la isla CpG asociada al gen PICALM.

Se realizó el análisis de metilación de estas regiones previamente a determinar si se producía la unión de ORC a las mismas. Para ello se utilizó, nuevamente, la técnica del bisulfito sódico, seguida de la amplificación con oligonucleótidos específicos sobre el DNA transformado y la secuenciación de los productos de PCR. Este análisis se realizó tanto para DNA de células 293 como para DNA de células mononucleadas de sangre y en ambos casos se pudo observar que las cinco regiones conservadas no codificantes establecían el límite de metilación de las islas asociadas en su extremo 5' (**Fig. 13**).

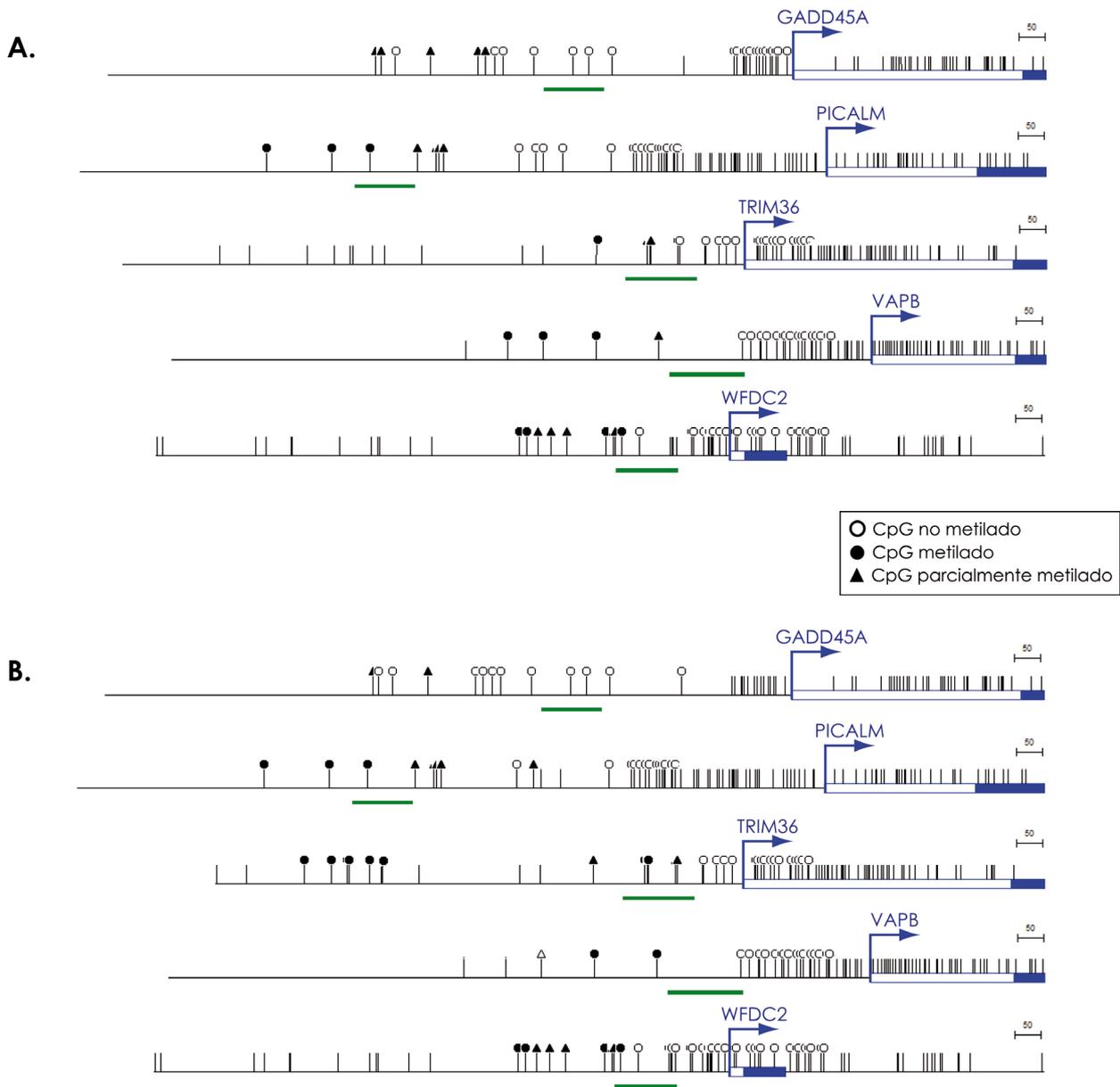


Figura 13. Análisis de metilación de las regiones conservadas no codificantes en el extremo 5' de las islas CpG asociadas a los genes *GADD45A*, *PICALM*, *TRIM36*, *VAPB* y *WFDC2*. A. Mapas de metilación en DNA de células mononucleadas de sangre periférica. **B.** Mapas de metilación en DNA de células 293.

Para determinar si las regiones seleccionadas unían ORC se utilizó la inmunoprecipitación de cromatina con anticuerpos específicos contra dos proteínas implicadas en el inicio de la replicación: ORC2 y CDC6. ORC2 es una de las seis subunidades que componen el complejo de reconocimiento del origen (ORC). El motivo de seleccionar la subunidad ORC2 es que se trata de la única subunidad del complejo ORC que permanece unida a los ORIs durante todo el ciclo celular en células humanas (Natale *et al.*, 2000), lo que facilitará su detección en poblaciones de células asincrónicas. CDC6, por su parte, forma parte del complejo pre-replicative (preRC), junto con ORC, CDT1 y el complejo de proteínas MCM (*Mini-Chromosome Maintenance*). CDC6 se une a los orígenes de replicación tras la unión del complejo ORC y es esencial para el reclutamiento de las proteínas MCM, que forman un anillo hexamérico con actividad helicasa (Borlado & Méndez, 2008) y son esenciales para el inicio de la replicación. Por tanto, CDC6 se une sólo a aquellos ORIs que se activan en cada ciclo celular.

Se realizó la inmunoprecipitación de cromatina a partir de células 293 creciendo en cultivo asincrónico. A continuación se analizó cada inmunoprecipitado mediante PCR cuantitativa con oligonucleótidos específicos para la región del ORI *LMNB2*, así como dos regiones flanqueantes situadas a 2,5 y 2,1 Kb a cada lado del ORI, lo que nos permite estimar el enriquecimiento de la región del ORI con respecto a los flancos. Los resultados se expresaron como el porcentaje del DNA de partida (*input*) que se ha inmunoprecipitado con cada anticuerpo para cada región analizada (**Fig. 14**).

Como se observa en la figura 14C, el porcentaje de *input* que se inmunoprecipitó con los anticuerpos contra ORC2 y CDC6 era muy inferior al detectado para la inmunoprecipitación de la histona H3, que se utilizó como control positivo, estando más próximo al valor del control negativo. Tras modificar las condiciones de la inmunoprecipitación y los anticuerpos utilizados en varias ocasiones éste es el mayor rendimiento que se logró obtener. A pesar de la baja eficiencia de la inmunoprecipitación se pudo detectar un enriquecimiento de la región del ORI con respecto a los flancos (**Fig. 14C**). Este enriquecimiento variaba en función de la proteína inmunoprecipitada, siendo de en torno a 4 veces para ORC2 y 2 veces para CDC6.

Para comprobar si se detectaba enriquecimiento en otro ORI se analizó el ORI *TOP1* con oligonucleótidos específicos para la región del ORI, así como para una región flanqueante situada a 0,9 Kb de éste. Nuevamente el porcentaje de *input* recuperado era muy bajo para ORC2 y CDC6 en comparación con H3, pero el enriquecimiento del ORI con respecto al flanco era similar al detectado para el ORI *LMNB2* (**Fig. 14D**).

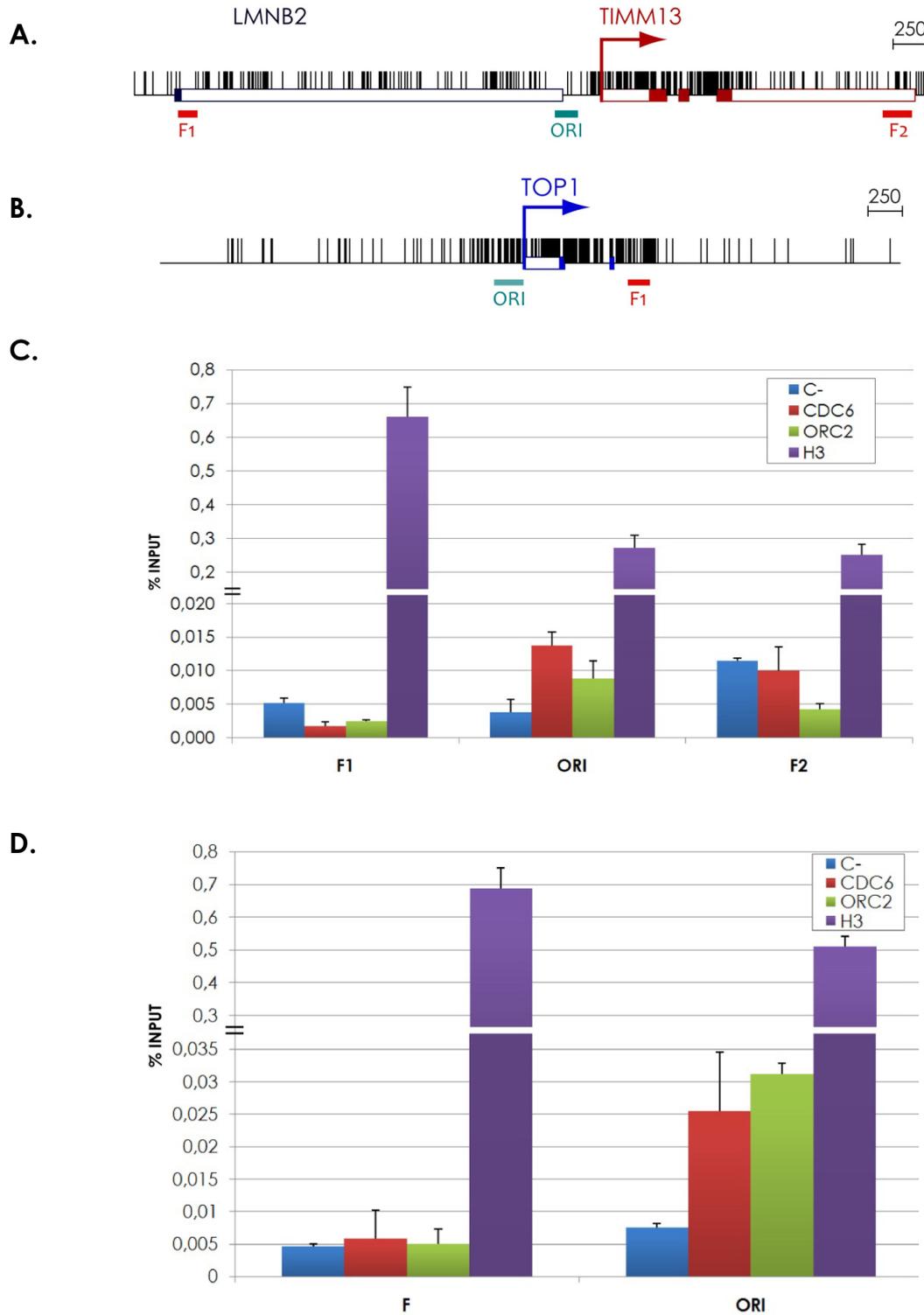


Figura 14. Análisis de la unión de ORC2 y CDC6 a los ORIS *LMNB2* y *TOP1* mediante inmunoprecipitación de cromatina. A. Mapa del locus del ORI *LMNB2* en el que se señalan las regiones amplificadas correspondientes al ORI y los flancos (F1 y F2). **B.** Mapa del locus del ORI *TOP1* en el que se señalan las regiones amplificadas correspondientes al ORI y el flanco (F). **C.** Nivel de unión de las distintas proteínas analizadas (H3, ORC2 y CDC6) al locus del ORI *LMNB2* expresado como porcentaje de INPUT recuperado. Como control negativo (C-) se incluyó en la inmunoprecipitación un tubo sin anticuerpo. **D.** Nivel de unión de las distintas proteínas analizadas (H3, ORC2 y CDC6) al locus del ORI *TOP1* expresado como porcentaje de INPUT recuperado.

A continuación, y a pesar del bajo rendimiento de la técnica, decidimos analizar en detalle tres de las regiones conservadas en el extremo 5' de islas CpG que habíamos seleccionado. En concreto, se analizaron las regiones del promotor de los genes *GADD45A*, *TRIM36* y *PICALM*. Se utilizaron para cada región varias parejas de oligonucleótidos con el fin de determinar el patrón de unión de ORC2 y CDC6 a lo largo de estas regiones. En el caso de *PICALM* (Fig. 15A), la máxima unión de ORC2 y CDC6 parecía coincidir con la región conservada (C) en el extremo 5' de la isla CpG, pero no ocurría lo mismo para *GADD45A* y *TRIM36*, donde ORC2 y CDC6 parecían unirse preferentemente en el centro de la isla (Fig. 15B y 15C). Estos resultados muestran que la región conservada no une la maquinaria de inicio de la replicación en dos de los tres casos analizados, y por tanto es probable que no funcionen como ORI. En el caso de *PICALM*, como se observa en el mapa de la figura 15A, no se pudo analizar la zona de la isla CpG ya que todas las parejas de oligonucleótidos que se probaron resultaron ser muy inespecíficas, de modo que no puede descartarse que la unión de ORC2 y CDC6 sea mayor en el interior de la isla CpG, como ocurre en el caso de *TRIM36* y *GADD45A*.

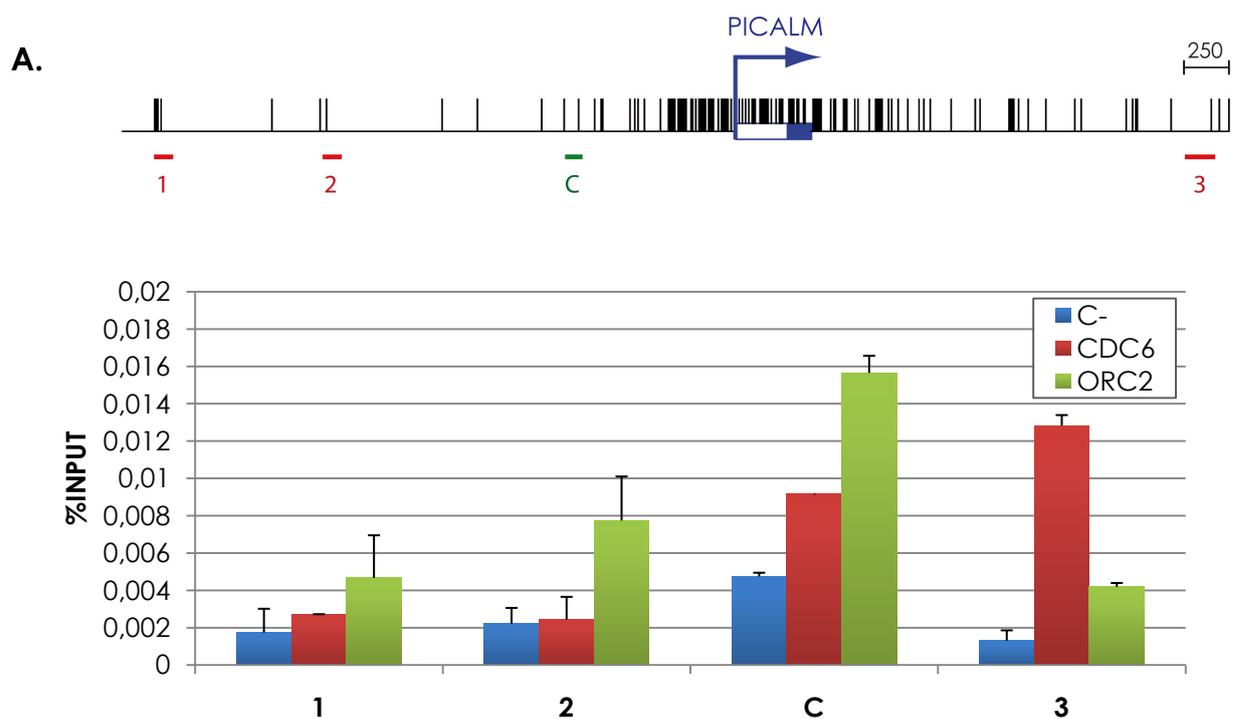


Figura 15. Análisis de la unión de ORC2 y CDC6 en las islas CpG asociadas a los genes *PICALM* (A), *GADD45A* (B) y *TRIM36* (C). Se indica con una C la posición de la región conservada.

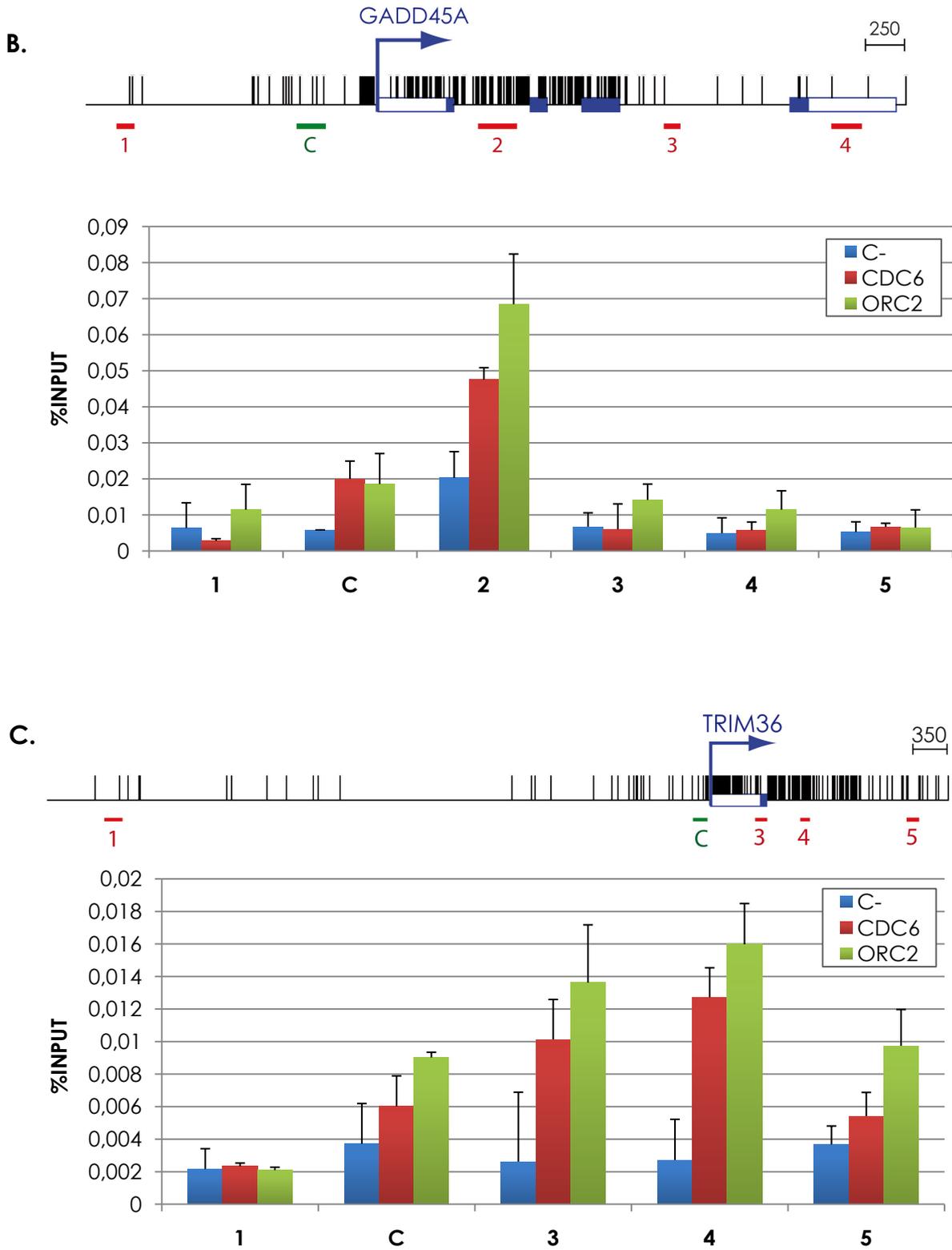


Figura 15. Análisis de la unión de ORC2 y CDC6 en las islas CpG asociadas a los genes *PICALM* (A), *GADD45A* (B) y *TRIM36* (C). Se indica con una C la posición de la región conservada.

Debido a la baja eficiencia de la inmunoprecipitación de cromatina decidimos comprobar que los resultados obtenidos eran fiables mediante el análisis de intermediarios de replicación en cadenas nacientes de DNA. Esta técnica permite determinar si una región está funcionando activamente como ORI en la población de células analizada.

Se extrajo el DNA de células 293 creciendo en cultivo asincrónico y se separó por tamaño utilizando gradientes neutros de sacarosa (Ver Materiales y Métodos). A continuación se analizó por PCR cuantitativa la presencia de las regiones problema en las tres primeras fracciones del gradiente, cuyos rangos de tamaños son 100-600 pb, 300-800 pb y 600-2000 pb. De este modo sólo las regiones que funcionen como ORIs deberían estar enriquecidas en la primera fracción (100-600 pb) mientras que las regiones flanqueantes comenzarían a mostrar enriquecimiento en fracciones que incluyan moléculas más grandes de DNA. Se analizó como control positivo el ORI *CDKN2B* utilizando una pareja de oligonucleótidos para la región del ORI y otra para una región flanqueante situada a 2 Kb. Como se observa en la figura 16 la región del ORI estaba enriquecida con respecto al flanco en la primera fracción y el enriquecimiento relativo disminuía en la segunda fracción de DNA analizada.

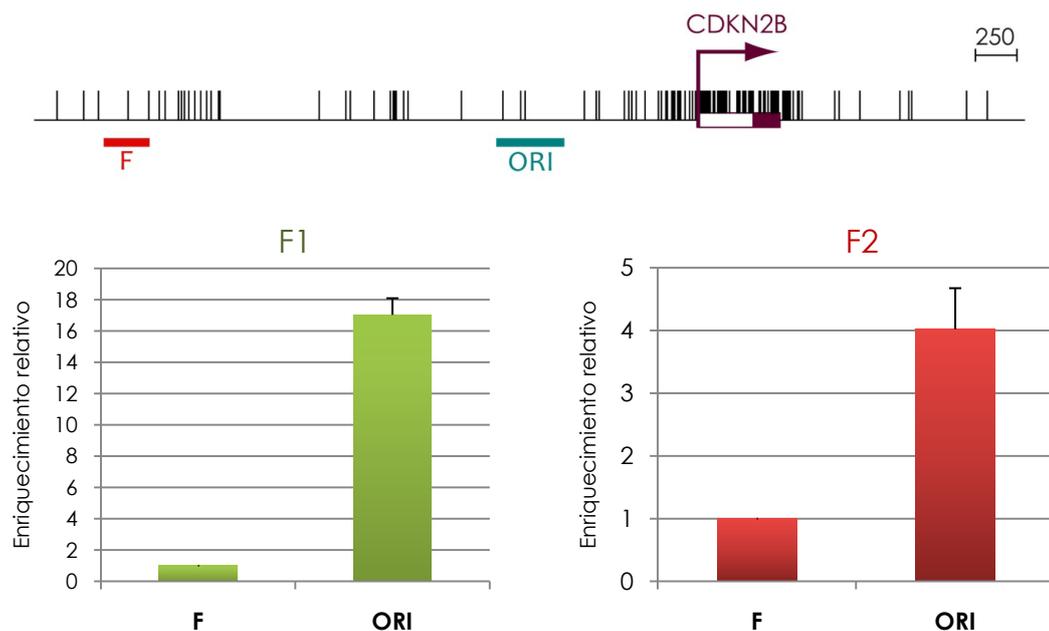


Figura 16. Análisis de intermediarios de replicación en cadenas nacientes de DNA para la región del ORI *CDKN2B*. A. Análisis de la región del ORI *CDKN2B* en las fracciones 1 y 2 del gradiente. Se representa el enriquecimiento relativo de la región del ORI con respecto a una región flanqueante (F). En el mapa se encuentran indicadas las regiones amplificadas.

A continuación se analizó la región de la isla asociada al gen *TRIM36*. Los resultados obtenidos mostraron que el mayor enriquecimiento se detectaba en la región central de la isla. Esto indicaba que, como sugerían los datos de inmunoprecipitación de *ORC2* y *CDC6*, la

replicación se iniciaba en el interior de la isla CpG y no en la región conservada que se encuentra en el extremo 5' de la misma (Fig. 17).

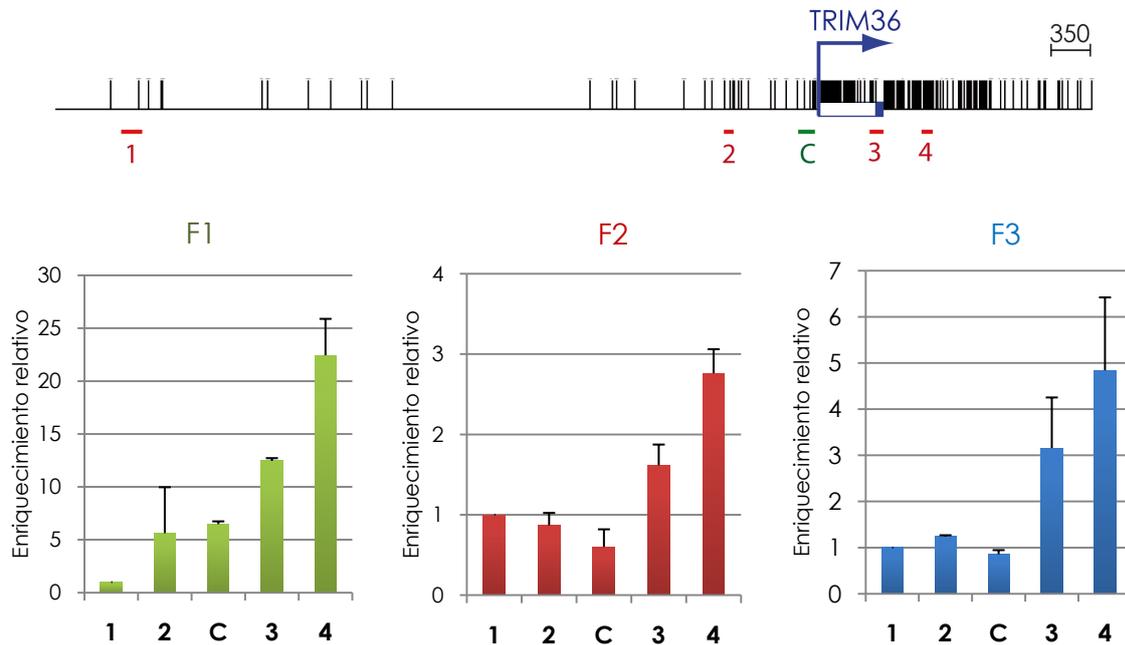


Figura 17. Análisis de intermediarios de replicación en cadenas nacientes de DNA para la región de la isla CpG asociada al gen TRIM36. Análisis de la isla CpG asociada al gen TRIM36 en las fracciones 1, 2 y 3 del gradiente. Se representa el enriquecimiento relativo con respecto a la región 1, que se ha considerado como flanco al ser la que menor abundancia presenta en la fracción 1 del gradiente.

Por tanto, con los datos obtenidos hasta el momento no podemos afirmar que sea el inicio de la replicación o la unión de los complejos proteicos responsables de ésta lo que establezca el límite de metilación de las islas en su extremo 5' puesto que en dos de los casos analizados, *TRIM36* y *GADD45A*, no existe una colocación entre inicio de la replicación y límite de metilación. Por otra parte, tampoco podemos afirmar que las regiones conservadas sean un buen indicador de la presencia de un ORI puesto que, a pesar de que parece iniciarse la replicación en todas las regiones analizadas, no existe una colocación espacial como la que tiene lugar en los ORIs *LMNB2* o *CDKN2B*.

Al hacer un cómputo de los datos obtenidos de metilación y unión de ORC en las regiones conservadas en el extremo de las islas CpG (Tabla 3) pudo observarse que, pese a que el número de regiones analizadas es pequeño, la correlación entre el límite de metilación y el inicio de la replicación no es tan alta como se postulaba en un principio, lo que sugiere que, aunque es posible que el inicio de la replicación contribuya a mantener las islas CpG libres de metilación, no parece que sea el principal mecanismo responsable de esta protección.

REGIÓN CONSERVADA	LÍMITE DE METILACIÓN	UNIÓN DE ORC
LMNB2	SÍ	SÍ
CDKN2B	SÍ	SÍ
PICALM	SÍ	SÍ*
GADD45A	SÍ	NO
TRIM36	SÍ	NO
VAPB	SÍ	-
WFDC2	SÍ	-

Tabla 3. Propiedades de las regiones localizadas en el extremo 5' de islas CpG que se encuentran conservadas filogenéticamente.

El asterisco indica que en el caso de PICALM no ha podido determinarse que la región conservada sea la que presenta la mayor unión de ORC puesto que no se ha conseguido analizar la región central de la isla CpG.

2. CONTRIBUCIÓN DE LA TRANSCRIPCIÓN AL MANTENIMIENTO DE LAS ISLAS CpG EN ESTADO NO METILADO

2.1 Antecedentes

Como se menciona en la introducción, la hipótesis más aceptada actualmente acerca del mecanismo que protege a las islas CpG frente a la metilación apunta a la maquinaria de inicio de la transcripción así como a la estructura de la cromatina asociada a las islas CpG como principales responsables (Liu *et al.*, 2008). Durante mucho tiempo se ha aceptado que las islas CpG permanecen libres de metilación de manera constitutiva, independientemente de la expresión del gen asociado. Sin embargo, existen varias evidencias que sugieren que, si bien la metilación no parece constituir un mecanismo primario de regulación de la expresión génica, la transcripción activa podría contribuir a la protección de las islas frente a la metilación.

Por una parte se ha observado que las islas CpG asociadas a genes específicos de tejido tienden a metilarse en líneas celulares en cultivo (Antequera *et al.*, 1990), mientras que no ocurre lo mismo con islas asociadas a genes de expresión constitutiva. Este dato sugiere que las islas CpG son más sensibles a metilarse *de novo* cuando se encuentran asociadas a genes que no se están transcribiendo activamente. Por otra parte, la reciente generación de mapas de metilación a nivel genómico ha mostrado que existen islas que se encuentran constitutivamente metiladas en tejidos somáticos (Yamada *et al.*, 2004; Weber *et al.*, 2005; Eckhardt *et al.*, 2006; Weber *et al.*, 2007; Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Rauch *et al.*, 2009; Straussman *et al.*, 2009; Illingworth *et al.*, 2010; Maunakea *et al.*, 2010). Las islas que se metilan con mayor frecuencia son aquellas que no se asocian a ningún promotor anotado, lo que nuevamente sugiere que la transcripción podría contribuir a prevenir la metilación de las islas CpG o, expresado en otros términos, que la ausencia de transcripción activa hace a una isla CpG más susceptible de metilarse.

2.2 Análisis de metilación de islas asociadas a genes específicos de tejido en ausencia de transcripción activa

Con la intención de determinar si las islas asociadas a genes que no se están expresando son más propensas a metilarse en tejidos normales en condiciones fisiológicas decidimos analizar el patrón de metilación de tres islas asociadas a genes humanos en un tejido en el que esos genes no se expresan.

Los genes seleccionados fueron *HBA1* (α -globina), *MYOD1* y *MT3* (Fig. 18B). El gen *HBA1* se expresa de forma específica en células del linaje eritroide. En el caso de *MYOD1*, su expresión se restringe al músculo y *MT3* presenta expresión específica de cerebro. El DNA sobre el que se realizó el análisis de metilación fue obtenido a partir de células mononucleadas de sangre periférica, donde ninguno de los tres genes se expresa. Además se analizaron dos islas asociadas a genes de expresión constitutiva, *TIMM13* y *TOP1* (Fig. 18A), para determinar si se detectaban diferencias de metilación con respecto a las islas asociadas a genes de expresión específica de tejido.

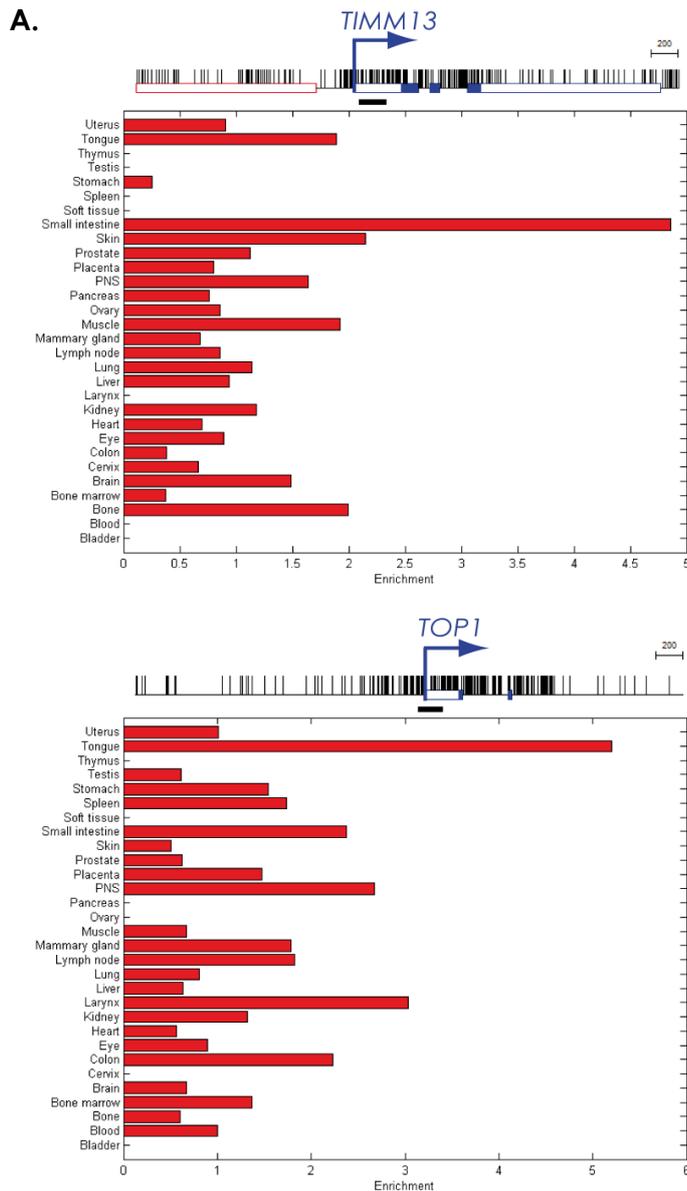


Figura 18. Selección de islas asociadas a genes de expresión constitutiva y genes de expresión específica de tejido. A. Perfiles de expresión de los genes de expresión constitutiva *TIMM13* y *TOP1* obtenida a partir de la base de datos TiGER (Tissue-specific Gene Expression and Regulation). En el eje X se representa el enriquecimiento en EST de un determinado tejido con respecto al valor esperado si el gen se expresara por igual en todos los tejidos. Se representa además, para cada gen, un mapa con la distribución de dinucleótidos CpG (líneas verticales), así como la región que se ha amplificado por PCR cuantitativa (barra negra).

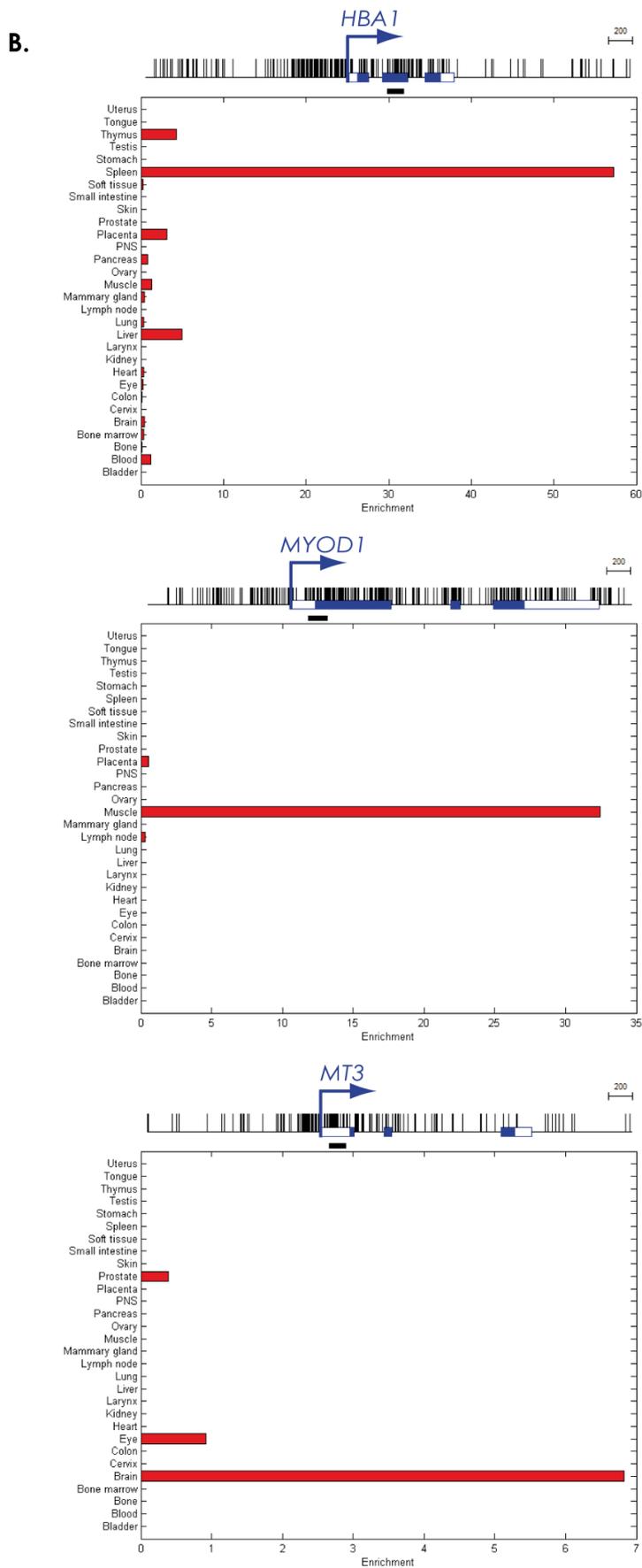


Figura 18. B. Perfiles de expresión de los genes de expresión específica de tejido *HBA1*, *MYOD1* y *MT3*.

Para llevar a cabo el análisis de metilación utilizamos la técnica de inmunoprecipitación de DNA metilado (MeDIP) (Weber *et al.*, 2005; Mohn *et al.*, 2009), que consiste en enriquecer en DNA metilado una determinada muestra de DNA mediante un anticuerpo que reconoce de forma específica las citosinas metiladas. A continuación pueden analizarse las regiones de interés mediante PCR cuantitativa. Esta técnica permite detectar de forma rápida niveles de metilación muy bajos, a diferencia de las técnicas clásicas de análisis de metilación como el bisulfito sódico o la digestión con enzimas de restricción sensibles a la metilación.

Sin embargo, esta técnica presenta una limitación importante y es que el enriquecimiento detectado para una región es directamente proporcional a la densidad de CpGs de la misma, lo que dificulta que se pueda estimar cuantitativamente el nivel de metilación de la región analizada (Fig. 19). Para evitar este sesgo, se diseñaron cuidadosamente las parejas de oligonucleótidos de modo que la región analizada en cada isla tuviera un tamaño y una densidad de CpGs similar y, de este modo, las señales de enriquecimiento obtenidas por MeDIP fueran comparables entre sí (Fig.18A).

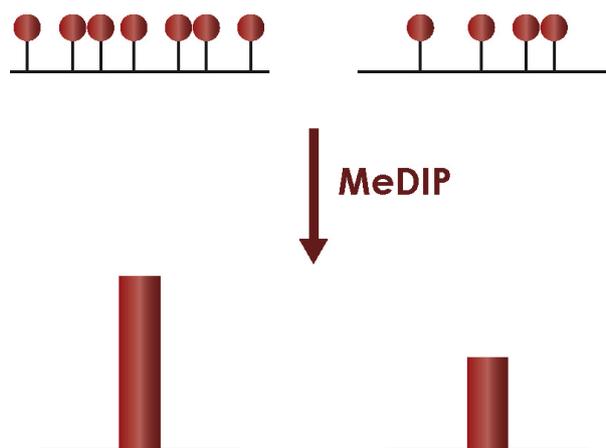


Figura 19. Efecto de la densidad de CpGs en la señal obtenida por inmunoprecipitación de DNA metilado. Dos regiones del mismo tamaño pero distinto contenido en CpGs presentarán una señal diferente en el inmunoprecipitado a pesar de estas ambas totalmente metiladas.

Siguiendo esta estrategia, se llevó a cabo el MeDIP sobre DNA de células mononucleadas de sangre periférica obtenido a partir de la combinación de tres muestras de sangre independientes. A continuación, se analizó el inmunoprecipitado por PCR cuantitativa con oligonucleótidos específicos para las regiones problema. Se incluyeron en el análisis una serie de regiones control, con niveles de metilación conocidos, para validar el funcionamiento correcto de la técnica. En primer lugar se analizó, como control negativo, una región que carece de CpGs y, por tanto, no es susceptible de metilarse. Además, como control positivo de metilación se incluyó una isla CpG que no se asocia a ningún promotor anotado y de la cual sabíamos, por experimentos previos de bisulfito, que se encuentra metilada en sangre. Finalmente se incluyeron dos controles correspondientes a regiones con niveles medios de metilación. Uno de ellos es *H19*, un gen sometido a *imprinting*, que tiene expresión monoalélica desde el alelo materno y que presenta la copia paterna metilada. También se incluyó como control de hemimetilación el gen *XIST*, que se encuentra en el cromosoma X y posee, en

hembras, un alelo metilado y otro no metilado en función de que se encuentre en el cromosoma X activo o inactivo, respectivamente. Es importante señalar que las tres muestras de sangre a partir de las cuales se realizó la extracción de DNA procedían de donantes femeninas.

La señal obtenida para cada región se normalizó con respecto al control negativo (NoCpG), de modo que lo que se representa en las figuras 20 - 23 es el enriquecimiento con respecto a una región no metilada.

Como se observa en la figura 20, el enriquecimiento detectado para *XIST* y *H19* era aproximadamente la mitad que para el control positivo, lo que indicaba que la técnica había funcionado correctamente. El nivel de metilación de las islas asociadas a genes de expresión constitutiva, *TIMM13* y *TOP1*, era similar al del control negativo. Esto significaba que esas islas estaban totalmente libres de metilación en DNA de sangre periférica. Sin embargo, las tres islas asociadas a genes de expresión específica de tejido presentaban un ligero enriquecimiento, lo que sugería que podrían presentar un cierto nivel de metilación en DNA de sangre, donde los genes asociados no se están expresando.

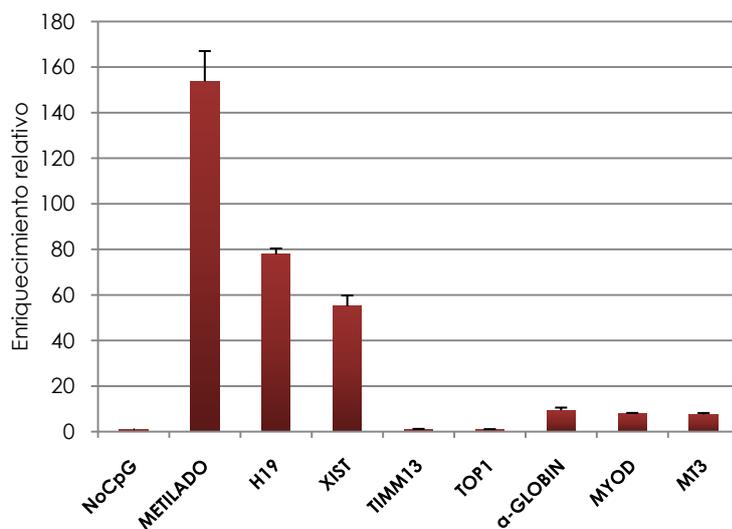


Figura 20. MeDIP sobre DNA de células mononucleadas de sangre periférica. Se representa el enriquecimiento relativo de cada región con respecto al control negativo de metilación (NoCpG). Se muestra el promedio de dos amplificaciones independientes y su desviación estándar para cada una de las regiones analizadas.

A continuación se llevó a cabo la inmunoprecipitación de DNA metilado sobre DNA procedente de un segundo conjunto de muestras de sangre (Fig. 21 S2), formado por tres muestras de sangre diferentes a las anteriores, que nuevamente se combinaron previamente a la extracción del DNA para realizar el análisis de metilación. En la figura 21 puede observarse que los resultados obtenidos eran similares a los obtenidos para el primer conjunto de muestras de sangre (Fig. 20 y Fig. 21 S1), salvo en el caso de la isla asociada al gen *MYOD1*, que presentaba un nivel de metilación superior al detectado en la primera inmunoprecipitación.

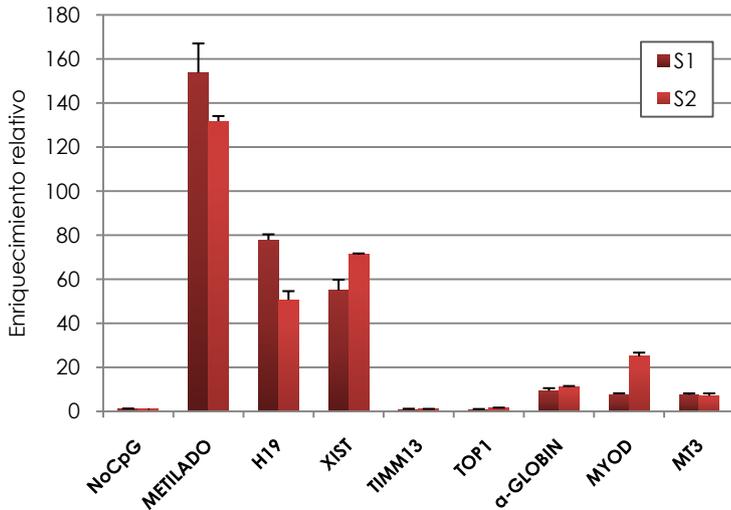


Figura 21. MeDIP sobre DNA de un segundo conjunto de muestras de sangre. Se representa el enriquecimiento relativo con respecto al control negativo de metilación (NoCpG) para cada una de las regiones analizadas en los dos conjuntos de muestras de sangres (S1 y S2) utilizados para el análisis. Se muestra, para cada región, el promedio de dos amplificaciones independientes y su desviación estándar.

Para comprobar cómo de homogéneos son los patrones de metilación entre las tres muestras de DNA de sangre que se combinaron para realizar el análisis de metilación, y descartar así que el enriquecimiento detectado fuera debido a la contribución de una sola muestra, realizamos la inmunoprecipitación de DNA metilado de manera independiente sobre el DNA de cada una de las tres muestras de sangre (S1.1, S1.2 y S1.3) que formaban parte del primer conjunto de sangres utilizado (Fig. 20). A continuación, analizamos por PCR cuantitativa varias de las regiones que se habían estudiado en el experimento anterior. Los resultados obtenidos (Fig. 22) demostraron que los patrones de metilación eran similares en las tres muestras analizadas y, por tanto, los niveles de metilación detectados en las islas asociadas a genes de expresión específica de tejido no podían atribuirse a un patrón de metilación presente únicamente en alguna de las muestras.

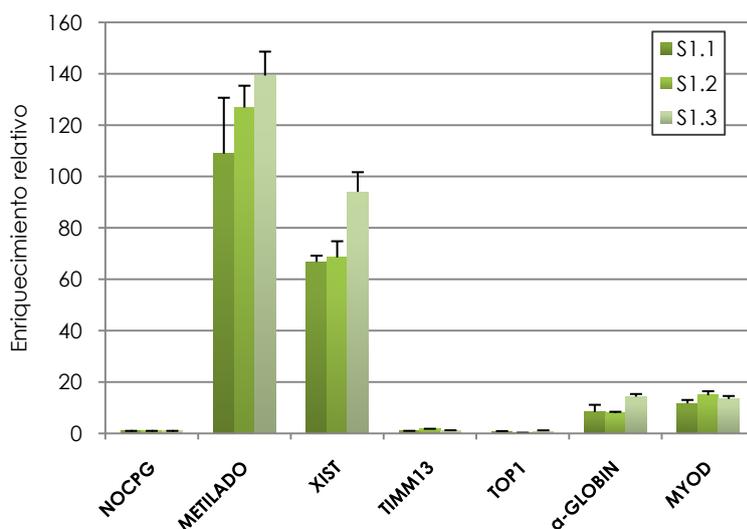


Figura 22. MeDIP sobre muestras independientes de DNA de sangre. Se ha representado el enriquecimiento relativo de cada región con respecto al control negativo de metilación (NoCpG) en cada una de las muestras analizadas (S1.1, S1.2 y S1.3). Se muestra promedio de dos amplificaciones independientes y su desviación estándar para cada una de las regiones analizadas.

Seguidamente, quisimos determinar si los niveles de metilación que mostraban las islas CpG asociadas a genes de expresión específica de tejido en DNA de sangre se detectaban también en la línea germinal. Por ello llevamos a cabo el MeDIP sobre DNA de esperma. El DNA de esperma presenta unos niveles globales de metilación inferiores a los detectados en tejidos somáticos y se ha observado que la gran mayoría de las islas CpG permanecen libres de metilación (Weber *et al.*, 2007; Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Straussman *et al.*, 2009). En este caso, *H19* representaba nuestro control positivo de metilación, ya que al tratarse de un gen de *imprinting* de expresión materna presenta el alelo paterno totalmente metilado. Como se observa en la figura 23, todas las demás islas, incluidas las asociadas a genes de expresión específica de tejido y la isla que habíamos utilizado como control positivo de metilación en DNA de sangre, mostraban un enriquecimiento similar al de la región desprovista de CpGs que utilizamos como control negativo. Estos resultados indicaban que la metilación detectada en las islas asociadas a los genes *HBA1*, *MYOD1* y *MT3* en DNA de sangre no tiene lugar en la línea germinal. Por otra parte, esto nos permitió afirmar que los niveles de metilación detectados en DNA de sangre eran significativos, descartando así que el enriquecimiento detectado para las islas asociadas a los genes *HBA1*, *MYOD1* y *MT3* con respecto a las islas de *TIMM13* y *TOP1* pudiera deberse al ruido de fondo de la inmunoprecipitación o a un sesgo por una amplificación más eficiente de estas regiones.

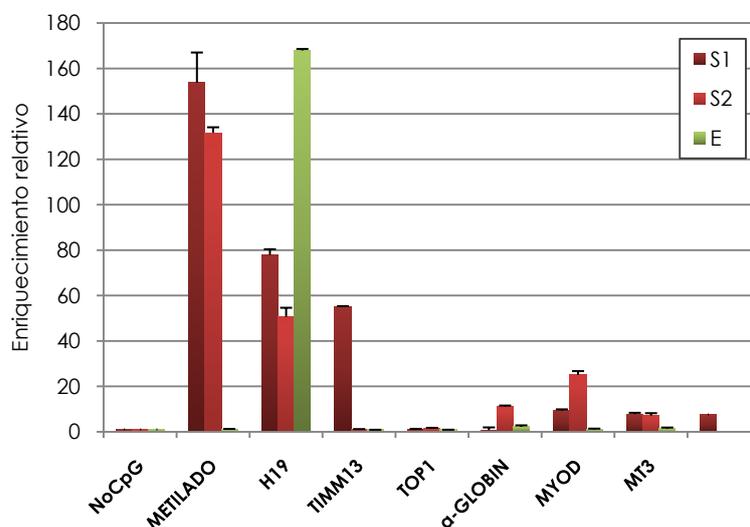


Figura 23. MeDIP comparativo sobre DNA de sangre y esperma. Se representa el enriquecimiento relativo con respecto al control negativo de metilación (NoCpG) para cada una de las regiones analizadas en los dos conjuntos de muestras de sangres (S1 y S2) y en el DNA de esperma (E). Se muestra el promedio de dos amplificaciones independientes y su desviación estándar para cada una de las regiones analizadas.

2.3 Análisis de metilación de la isla asociada al gen de la α -globina en condiciones de transcripción activa

A continuación, con el objetivo de averiguar si las diferencias observadas en el nivel de metilación en las islas CpG asociadas a genes de expresión específica de tejido eran atribuibles

a la ausencia de transcripción activa, decidimos analizar el estado de metilación de alguna de estas islas en un tejido en el que el gen asociado se expresara activamente. Ante la dificultad de obtener muestras de los tejidos humanos en los que se expresan estos genes (eritroblastos, músculo y cerebro) decidimos utilizar DNA de la línea celular K562, procedentes de leucemia mieloide crónica, en las que se expresa de forma activa el gen de la α -globina.

Al analizar por PCR cuantitativa el inmunoprecipitado no fuimos capaces de amplificar la región correspondiente al control negativo sin CpGs y tampoco pudimos amplificarla en el DNA genómico sonicado utilizado como material de partida para la inmunoprecipitación, lo que nos llevó a pensar que esta región habría sufrido alguna mutación o estaría delecionada en esta línea celular. Por tanto, al carecer de datos para esta región, utilizamos como control negativo para este experimento la isla asociada al gen TIMM13, de expresión constitutiva, que en todos los experimentos anteriores presentaba valores de metilación similares a los de la región desprovista de CpGs.

Los resultados obtenidos mostraron que la isla asociada al gen de la α -globina presentaba en células K562 un enriquecimiento similar al que se detectaba en esperma y similar al de la isla asociada al gen TIMM13, lo que indicaba que se encontraba desprovista de metilación en estas células. Por otra parte, se observó que tanto la isla asociada al gen de imprinting *H19*, como las otras dos islas asociadas a genes específicos de tejido, *MYOD1* y *MT3*, se encontraban totalmente metiladas en esta línea celular (**Fig. 24**).

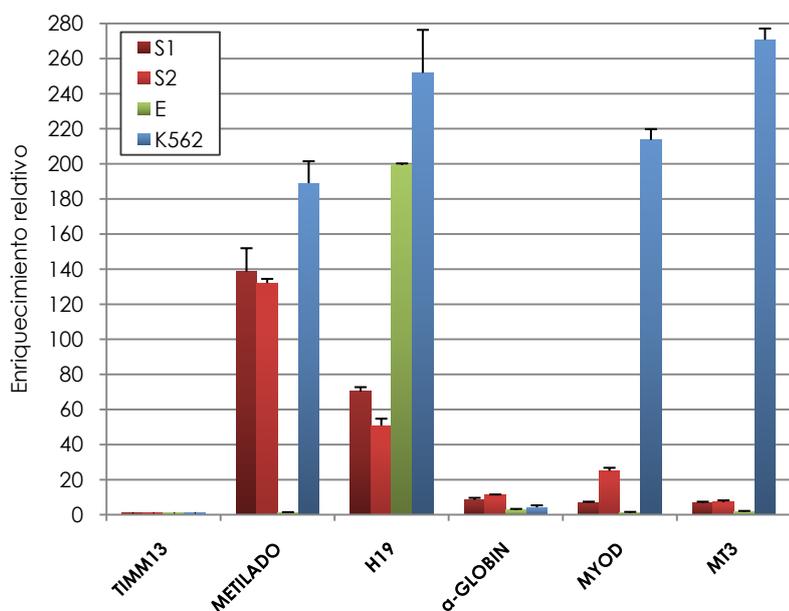


Figura 24. MeDIP sobre DNA de células K562. Se representa el enriquecimiento relativo con respecto al control negativo de metilación (NoCpG) para cada una de las regiones analizadas en los dos conjuntos de muestras de sangre (S1 y S2), esperma (E) y K562. Se muestra el promedio de dos amplificaciones independientes y su desviación estándar para las distintas regiones analizadas en cada muestra.

2.4 Análisis del patrón de metilación de la isla CpG asociada al gen de la α -globina en ausencia de transcripción activa

Los resultados obtenidos mediante los experimentos de MeDIP sobre DNA de sangre periférica, espermatozoides y células K562 mostraban que las islas asociadas a genes específicos de tejido presentaban un ligero incremento en el nivel de metilación en ausencia de transcripción activa. Sin embargo, el enriquecimiento detectado puede tener varias interpretaciones en lo referente al patrón de metilación. Este enriquecimiento podría deberse a que un bajo porcentaje de las células de la población tuvieran la región analizada densamente metilada, pero también se detectaría si todas las células de la población tuvieran unos pocos CpGs metilados en la región analizada. Para determinar cuál podría ser el patrón de metilación de estas regiones analizamos en más detalle la isla CpG asociada al gen de la α -globina. Para ello, diseñamos una segunda pareja de oligonucleótidos (**Fig. 25A "G1"**) para analizar la mitad de la isla CpG situada hacia el extremo 5' de la misma. La pareja de oligonucleótidos utilizada en los experimentos anteriores (**Fig. 25A "G2"**) amplificaba la mitad de la isla situada hacia el extremo 3'. Las regiones analizadas por cada pareja de oligonucleótidos poseían una densidad de CpGs muy similar de modo que las señales obtenidas mediante PCR cuantitativa eran comparables. Como control, diseñamos también una segunda pareja de oligonucleótidos (**Fig. 25A "T2"**) para la isla asociada al gen *TOP1*, como se muestra en la figura 25. A. Con estos nuevos oligonucleótidos analizamos el inmunoprecipitado de las tres muestras de sangre independientes utilizadas en experimentos anteriores (**Fig. 22 S1.1, S1.2 y S1.3**) y pudimos observar que en las tres el enriquecimiento detectado era mayor hacia el extremo 3' de la isla de la α -globina, mientras que la región correspondiente a la región 5' de la misma mostraba un nivel de metilación inferior. En el caso de la isla asociada al gen *TOP1* no se observaron diferencias entre las dos regiones analizadas, mostrando ambas un enriquecimiento similar al del control negativo.

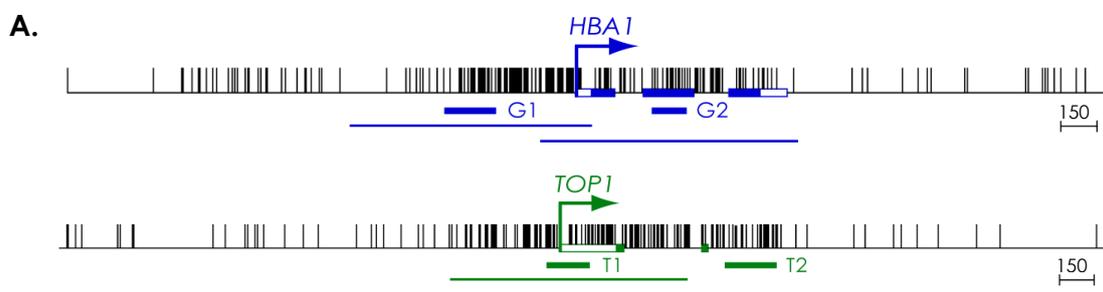


Figura 25. Análisis del patrón de metilación de la isla asociada al gen de la α -globina. A. Mapa de las islas CpG asociadas a los genes *HBA1* y *TOP1* en los que se indican las regiones amplificadas (línea gruesa) así como las regiones cuyo patrón de metilación influirá directamente en la señal del MeDIP teniendo en cuenta el tamaño promedio de los fragmentos de DNA utilizados (línea fina).

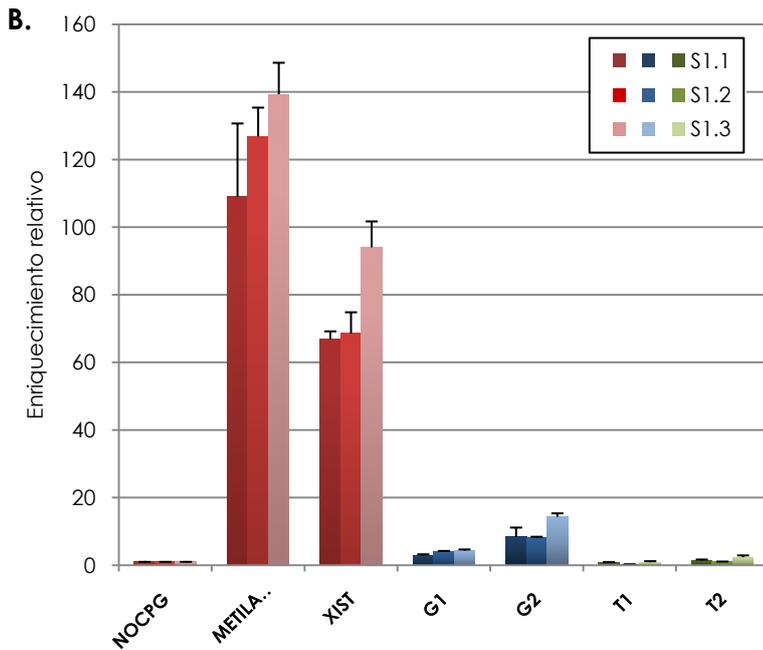


Figura 25. B. Se representa el enriquecimiento relativo de cada región con respecto al control negativo de metilación (NoCpG) en cada una de las tres muestras de sangre analizadas (S1.1, S1.2 y S1.3). Se muestra el promedio de dos amplificaciones independientes y su desviación estándar para cada una de las regiones analizadas.

Estos resultados parecen indicar que, en ausencia de transcripción activa del gen asociado, la isla de la α -globina podría estar comenzando a metilarse por el extremo 3', haciendo así que se reduzca progresivamente la región libre de metilación.

2.5 Análisis de metilación mediante MeDIP combinado con la hibridación de arrays de alta densidad (MeDIP-CHIP)

Para caracterizar a nivel genómico el fenómeno descrito en los apartados anteriores por el que las islas CpG asociadas a genes específicos de tejido son más sensibles a la metilación en ausencia de transcripción activa decidimos combinar la inmunoprecipitación de DNA metilado con la hibridación de arrays de alta densidad (*tiling microarrays*). Entre los 7 arrays que componen el *GeneChip® Human Tiling 2.0R Array Set* de Affymetrix, seleccionamos el que contiene los cromosomas 8, 11 y 12.

2.5.1 Diseño experimental

Como se menciona en el apartado 2.2, la técnica del MeDIP presenta un sesgo, que hemos denominado efecto de la densidad de CpGs, y que provoca que la señal de metilación obtenida no sea cuantitativa, sino directamente proporcional a la densidad de CpGs metilados en una región (**Fig. 19**). Para corregir este efecto al analizar el inmunoprecipitado por PCR cuantitativa, diseñamos los oligonucleótidos de tal manera que las regiones analizadas presentasen una densidad similar de CpGs y, por tanto, fuesen directamente comparables. Sin

embargo, esta estrategia no es aplicable a la hibridación de *tiling microarrays*. En este caso utilizamos una aproximación distinta para solventar el problema del efecto de la densidad de CpGs.

Esta nueva estrategia consistió en la obtención de una copia del genoma totalmente metilada mediante metilación *in vitro* con la enzima metilasa *SssI*. De este modo, al realizar MeDIP sobre esta muestra obtendríamos una señal directamente proporcional a la densidad de CpGs. Así, las islas CpG, al ser regiones con una alta densidad de CpGs, presentarían los mayores niveles de enriquecimiento, el resto del genoma mostrará niveles intermedios y las regiones desprovistas de CpG serían las que mostrarán los niveles más bajos (**Fig. 27** línea roja). En cambio, si el MeDIP se realizaba sobre una muestra de DNA no sometida a metilación *in vitro*, la señal que se obtendría dependería del nivel de metilación de cada región así como de la densidad de CpGs de la misma. Así, esperaríamos encontrar la mayoría del genoma altamente metilado, a excepción de las islas CpG, que se encuentran libres de metilación en su mayor parte (**Fig. 27** línea verde). Si normalizábamos la señal del inmunoprecipitado problema frente a la señal del inmunoprecipitado de DNA metilado obtendríamos un ratio de metilación diferencial (**Fig. 27** línea azul) en el que las regiones que se encontraran metiladas presentarían un valor próximo a 1, mientras que aquellas regiones desprovistas de metilación tendrían valores cercanos al 0.

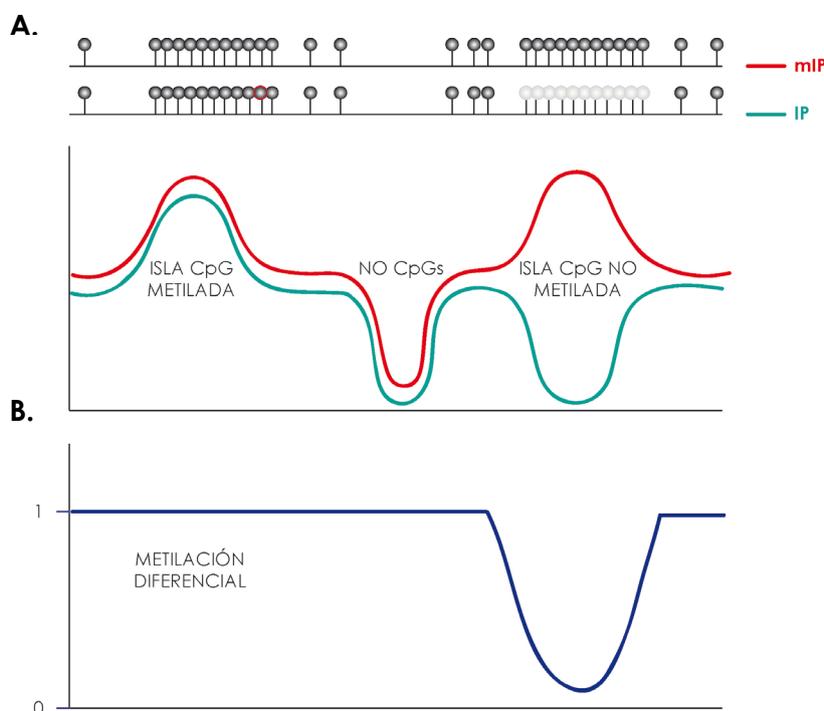


Figura 27. Estrategia de normalización de los datos de MeDIP-CHIP. **A.** Se representa el mapa de metilación y la señal que se espera detectar en el inmunoprecipitado de una muestra problema (IP; línea verde) y en una muestra sometida a metilación *in vitro* (mIP; línea roja). **B.** Representación del cociente entre la señal del inmunoprecipitado de la muestra problema y la señal del inmunoprecipitado de la muestra metilado (línea azul).

2.5.2 Metilación de DNA *in vitro*

En primer lugar se llevó a cabo la metilación *in vitro* de DNA de esperma, previamente sonificado, mediante la enzima *SssI*, que metila la citosina de todos los dinucleótidos CpG. En este caso el rango de tamaños utilizado oscilaba entre 100 y 500 pb, siendo ligeramente inferior al utilizado para los experimentos de MeDIP-PCR. El motivo de utilizar DNA de menor tamaño fue incrementar la resolución de la técnica. A continuación, se validó la correcta metilación del DNA. Para ello se transformó una parte del DNA metilado con bisulfito sódico y se analizó la isla CpG asociada al gen *TIMM13*, una región que, como se muestra en la figura 23, se encuentra libre de metilación en DNA de sangre y esperma. Los resultados mostraron que la isla se encontraba densamente metilada en la muestra de DNA metilado *in vitro* (Fig. 28).

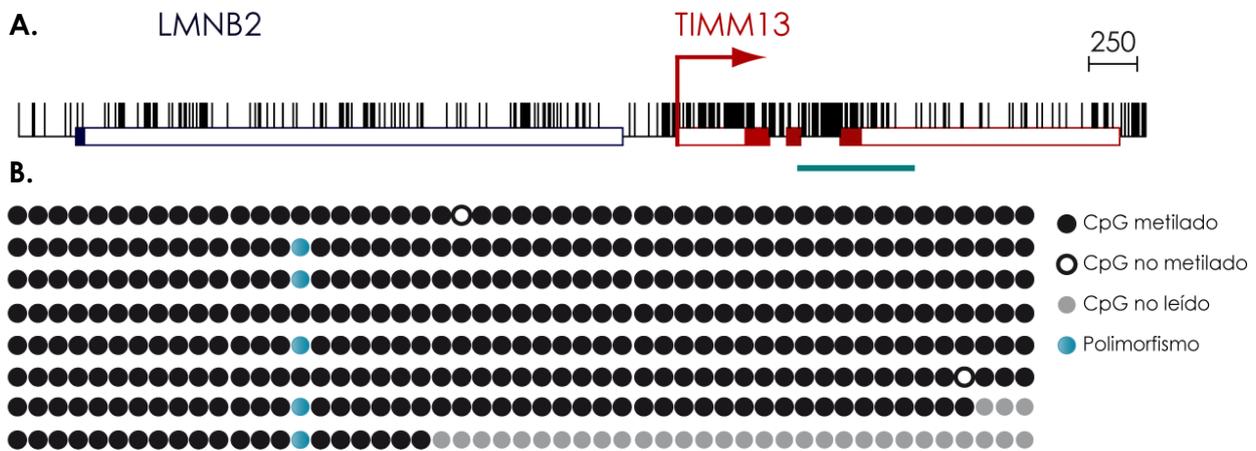


Figura 28. Validación de la reacción de metilación *in vitro* por bisulfito sódico. **A.** Mapa de la isla CpG asociada al gen *TIMM13* en el que se representa la región analizada por bisulfito sódico. **B.** Patrón de metilación de los distintos clones analizados.

2.5.3 MeDIP-PCR

Seguidamente se llevó a cabo la inmunoprecipitación de DNA metilado sobre DNA de sangre periférica, DNA de esperma y DNA metilado de esperma. Para cada muestra de DNA se realizó un duplicado técnico. Se realizó la validación del MeDIP por PCR cuantitativa analizando un control negativo de metilación (NoCpG), un control positivo correspondiente a una isla CpG que se encuentra metilada en DNA de sangre (Metilado) y el gen de *imprinting* H19 como control de hemimetilación. Como se observa en la figura 29 se obtuvieron en cada caso los resultados esperados, confirmando que la técnica había funcionado correctamente.

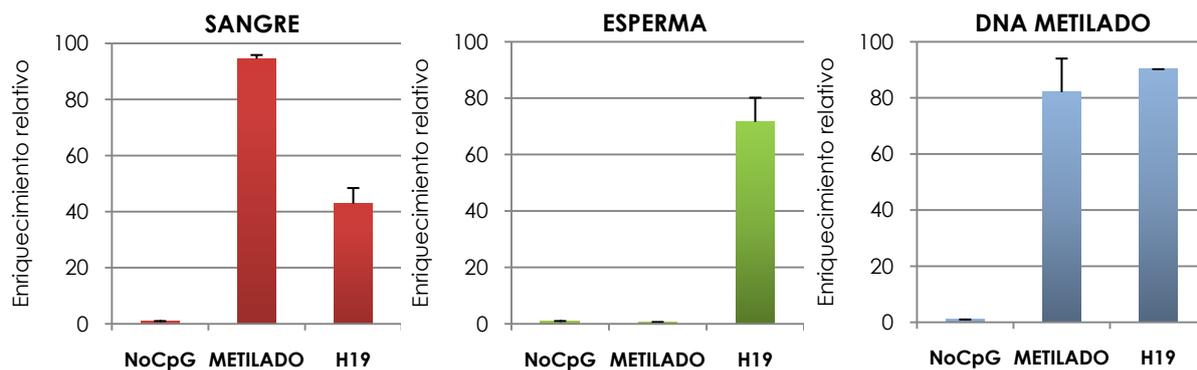


Figura 29. Análisis mediante PCR cuantitativa de los inmunoprecipitados de DNA de sangre, DNA de esperma y DNA metilado de esperma. Se muestra el promedio de dos inmunoprecipitaciones independientes y su desviación estándar para cada región analizada.

2.5.4 Amplificación de DNA inmunoprecipitado

Se estima que en una reacción de MeDIP en la que se parte de 4 µg de DNA se inmunoprecipita un 5% de esta cantidad, es decir, unos 200 ng. Para hibridar un microarray se necesitan, al menos, 5 µg de DNA, por lo que es necesario amplificar el DNA inmunoprecipitado como paso previo a la hibridación. Para evitar que la amplificación ocasionara la pérdida del enriquecimiento relativo obtenido mediante la inmunoprecipitación, se utilizó el menor número posible de ciclos de PCR. Con este objetivo se tomaron dos alícuotas de cada inmunoprecipitado y se amplificaron tal y como se indica en el apartado de materiales y métodos. A continuación, se combinaron las alícuotas de cada inmunoprecipitado y se analizó una dilución 1/20 del producto de la amplificación por PCR cuantitativa. Los resultados obtenidos (Fig. 30) mostraron que, tras el proceso de amplificación, se mantenía el enriquecimiento relativo entre nuestros controles de metilación, hemimetilación (H19) y ausencia de metilación (NoCpG).

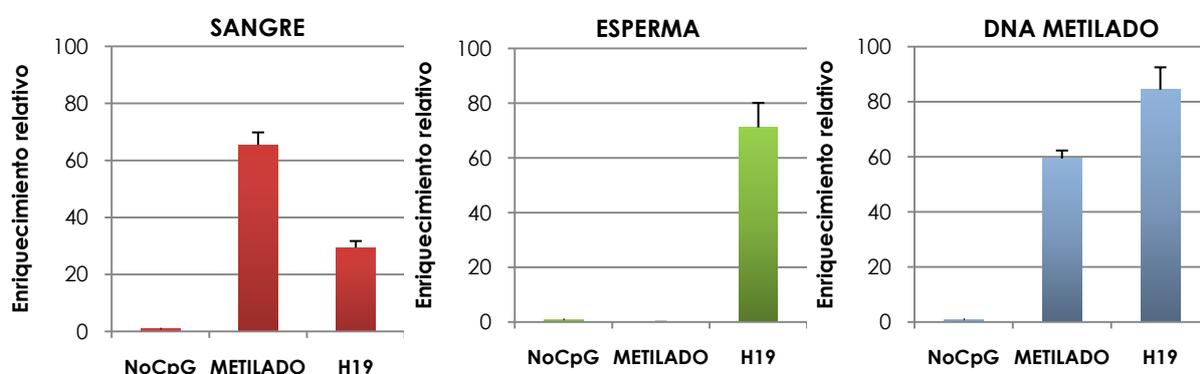


Figura 30. Análisis mediante PCR cuantitativa de los inmunoprecipitados de DNA de sangre, DNA de esperma y DNA metilado tras la amplificación. Se representa el promedio de dos amplificaciones independientes y su desviación estándar para cada una de las regiones analizadas.

2.5.5 Hibridación de *tiling microarrays*

Una vez se hubo comprobado que se mantenía el enriquecimiento relativo tras la amplificación, se cuantificaron las muestras y se enviaron 7 µg de cada una a la Unidad de Genómica de la Universidad de Salamanca, donde se llevó a cabo la hibridación de los *microarrays*. Los datos obtenidos a partir de la hibridación de los *tiling microarrays* se normalizaron utilizando la estrategia descrita en la figura 27 mediante el programa TAS (*Tiling Analysis Software*) de Affymetrix y a continuación se representaron en el navegador IGB (*Integrated Genome Browser*) (Nicol *et al.*, 2009).

2.5.6 Análisis de los datos de MeDIP-CHIP

Para analizar los resultados se estableció un umbral por debajo del cual consideraríamos a una región como no metilada. Así, se establecieron como regiones no metiladas todas aquellas que presentasen un valor inferior a -0,3 para el logaritmo en base 2 del cociente entre el valor del MeDIP sobre DNA de sangre o esperma y el valor del MeDIP sobre DNA metilado *in vitro* ($\log_2 (IP/mIP) \leq -0,3$). Además se estableció un tamaño mínimo de 400 pb. Utilizando este umbral la mayor parte de las regiones que se detectaban como no metiladas correspondían a islas CpG (Fig. 31). Se utilizó el listado de islas CpG del navegador UCSC Browser (Kent *et al.*, 2002) que utiliza el algoritmo de Gardiner-Garden and Frommer (Gardiner-Garden & Frommer, 1987).

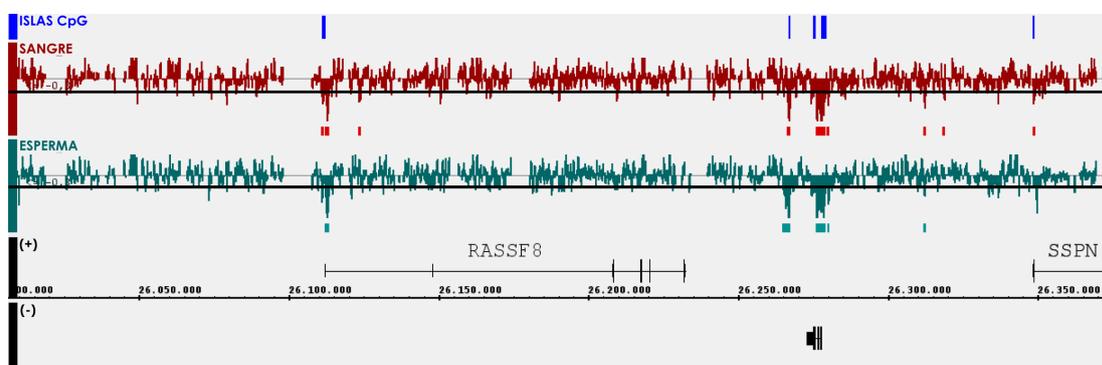


Figura 31. Colocalización de islas CpG y regiones detectadas como no metiladas por MeDIP-CHIP. Se muestra una región del cromosoma 12 en la que se representan las islas CpG (barras azules en la parte superior) y los datos de metilación del DNA de sangre (rojo) y esperma (verde). Se utilizó el listado de islas CpG del navegador UCSC Browser que utiliza el algoritmo de Gardiner-Garden and Frommer (Gardiner-Garden & Frommer, 1987). Además se representan los genes presentes en esta región, tanto en la cadena positiva (+) como en la cadena negativa (-). Las regiones que presentaban un valor de $\log_2 (IP/mIP) \leq -0,3$ a lo largo de al menos 400 pb se consideraron como no metiladas y se señalan en la figura con recuadros situados debajo de cada histograma, que son de color rojo en el caso del DNA de sangre y de color verde en el caso del DNA de esperma.

Mediante este criterio pudimos observar que se detectaban como no metiladas en torno a un 40% de las islas CpG contenidas en los tres cromosomas analizados. Estableciendo un umbral más estricto se reducía el número de regiones no clasificadas como islas, pero también el número de islas CpG que se detectaban como no metiladas. En cambio, si se establecía un umbral más flexible eran más las islas que aparecían como no metiladas, pero también se detectaba un mayor número de regiones no metiladas que no se clasificaban como islas CpG.

2.5.7 Validación de los resultados por MeDIP-PCR y análisis de metilación por bisulfito sódico

A continuación quisimos determinar cómo de fiable era la información obtenida validando algunas regiones mediante PCR cuantitativa sobre el inmunoprecipitado y, en algunos casos, también mediante análisis de metilación por bisulfito sódico. En primer lugar, llevamos a cabo el análisis de islas CpG que se detectaban como metiladas por presentar valores superiores al umbral establecido. Se seleccionaron dos islas CpG, asociadas a los promotores de los genes *TRIM28* Y *GRIN2B*, y se diseñaron oligonucleótidos para su análisis por PCR cuantitativa en el inmunoprecipitado de sangre. Como puede observarse en la figura 32B, ambas mostraron un enriquecimiento similar a nuestro control positivo de metilación.



Figura 32. Islas metiladas en DNA de sangre. A. Datos de MeDIP-CHIP para las islas asociadas a los genes *TRIM29* y *GRIN2B* en DNA de sangre.

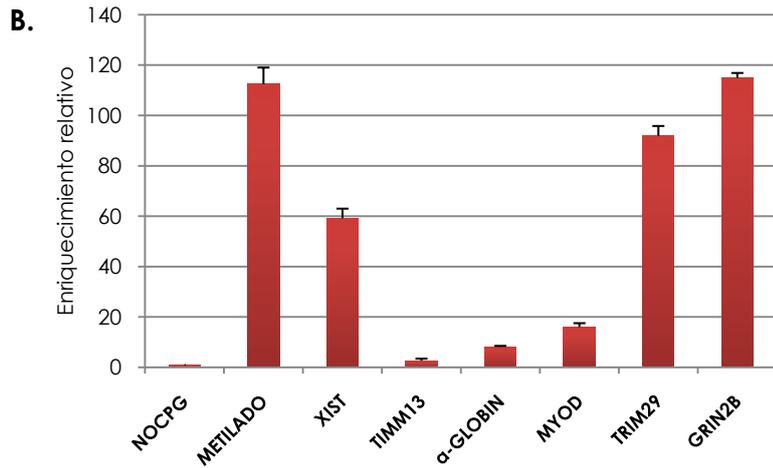


Figura 32. B. Validación del estado de metilación de estas islas por MeDIP-PCR.

A continuación se seleccionó una isla diferencialmente metilada entre sangre y esperma (**Fig. 33A**) y se analizó por PCR cuantitativa en ambos inmunoprecipitados, confirmándose las diferencias observadas en el *microarray* (**Fig. 33C**). Además, en este caso se confirmó mediante análisis de metilación por bisulfito sódico que la isla analizada estaba metilada en DNA de sangre (**Fig. 33B**).

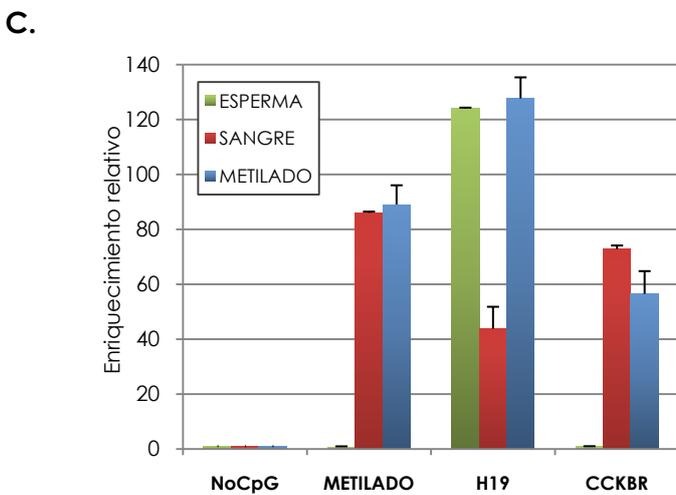
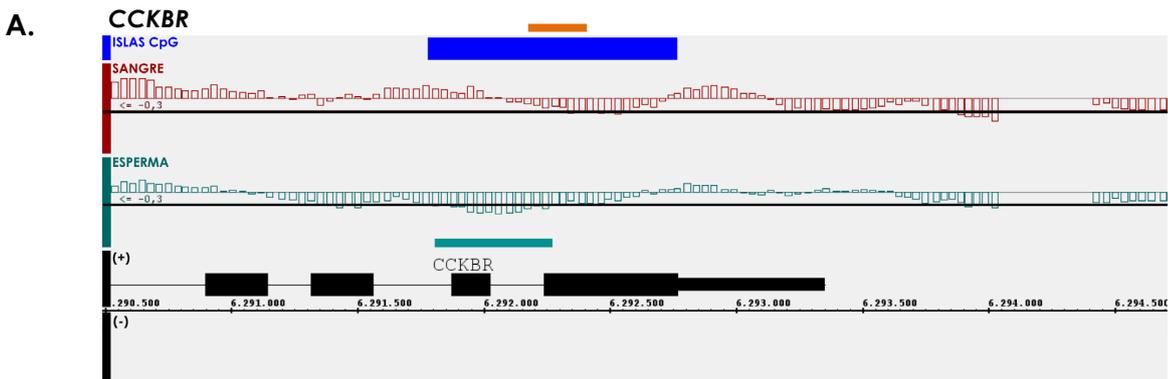


Figura 33. Isla CpG diferencialmente metilada entre sangre y esperma. A. Datos de MeDIP-CHIP para la isla CpG asociada al último exón del gen CCKBR en DNA de sangre y esperma. Se indica con una línea naranja la región que ha sido analizada por bisulfito sódico **B.** Análisis de metilación por bisulfito sódico de esta región en DNA de sangre. Los círculos negros indican aquellos dinucleótidos CpG que se encuentran metilados. **C.** Validación del estado de metilación de la isla por MeDIP-PCR.

Finalmente, con la intención de determinar si las regiones detectadas como no metiladas que no correspondían a islas CpG estaban realmente desprovistas de metilación, se seleccionaron varias de estas regiones. Escogimos una región no metilada en sangre pero sí en esperma (NIHS: No Isla Hipometilada en Sangre), una región metilada en sangre pero no en esperma (NIHE: No Isla Hipometilada en Esperma) y una región no metilada en esperma ni en sangre (NIHE+S: No Isla Hipometilada en Esperma y Sangre) (**Fig. 34A**). Nuevamente, el análisis por MeDIP-PCR confirmó los datos observados en el *microarray* (**Fig. 34B**). Se realizaron además varias validaciones por bisulfito sódico, que confirmaron los datos obtenidos (**Fig. 34C**).

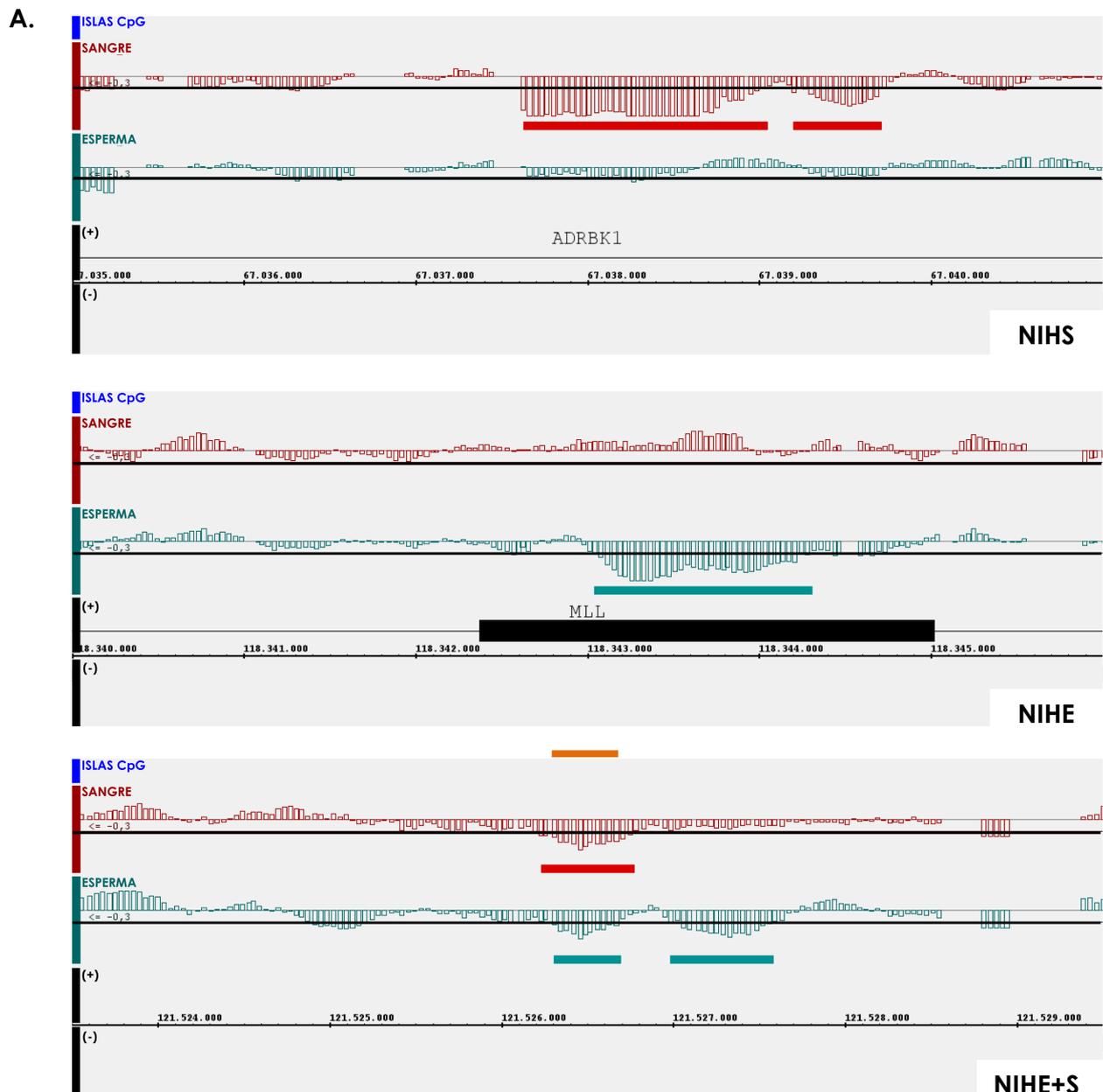
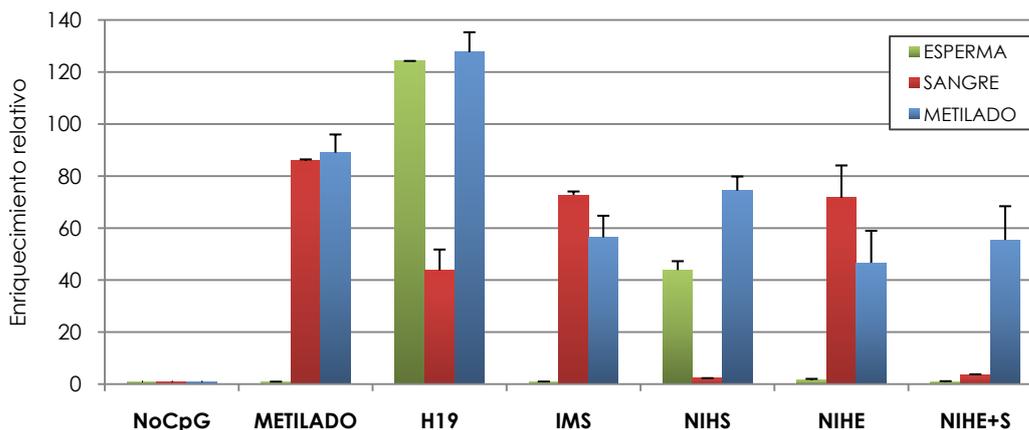


Figura 34. Regiones no metiladas que no son islas CpG. A. Datos de MeDIP-CHIP para varias regiones no metiladas y no clasificadas como islas CpG: NIHS (No Isla Hipometilada en Sangre), NIHE (No Isla Hipometilada en Esperma) y NIHE+S (No Isla Hipometilada en Esperma y Sangre). La línea naranja indica la zona analizada por bisulfito sódico para la región no metilada en esperma y sangre (NIHE+S).

B.



C.

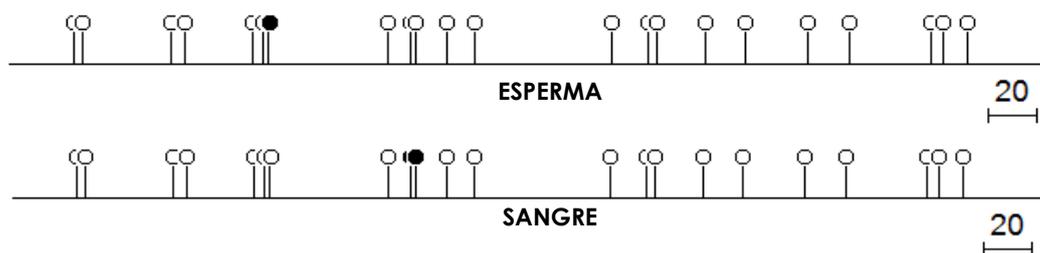


Figura 34. B. Validación del estado de metilación de estas regiones por MeDIP- PCR. Se han incluido los datos obtenidos para la isla asociada al gen *CCKBR* (IMS: Isla Metilada en Sangre) obtenidos en el experimento anterior. **C.** Análisis de metilación por bisulfito sódico de la región NIHE+S sobre DNA de esperma y DNA de sangre.

Todas las validaciones realizadas hasta este punto indicaban que la combinación del MeDIP con la hibridación de *tiling microarrays* es una buena herramienta para detectar islas CpG metiladas en el genoma y nos permite, además, detectar regiones no clasificadas como islas pero que se encuentran desprovistas de metilación.

2.5.8 Análisis comparativo de islas CpG con distintos niveles de metilación

Seguidamente quisimos comprobar si la técnica de MeDIP-CHIP era capaz de detectar diferencias más pequeñas en el nivel de metilación de distintas islas CpG. Para ello decidimos analizar el patrón detectado para una isla asociada a un gen de expresión constitutiva, *MCM4* y una isla asociada a un gen de expresión específica de tejido, *MYOD1*. El motivo de utilizar la isla asociada al gen *MCM4* en lugar de las islas de *TIMM13* o *TOP1* es que ninguna de estas dos se encuentra representada en el *microarray*, puesto que este incluye únicamente los cromosomas 8, 11 y 12.

Según los datos obtenidos en el apartado 2.2 (Fig. 20 y 21, Pág. 56), la isla de *MYOD1* tendría que presentar su isla ligeramente metilada en sangre. Como puede observarse en la figura 35, se apreciaba una ligera diferencia en la isla del *MYOD1* entre sangre y esperma, lo que coincide con los datos de MeDIP-PCR. Sin embargo, estas diferencias eran similares a las que se podían observar en la isla asociada al gen *MCM4*, de expresión constitutiva, en el que se esperaba, por tanto, que la isla no estuviera metilada ni en sangre ni en esperma. Esto indica que, o bien la isla asociada a *MCM4* se encontraba ligeramente metilada en sangre, o nos encontrábamos en el límite de metilación de la técnica de modo que éramos capaces de distinguir diferencias tan sutiles en el nivel de metilación.

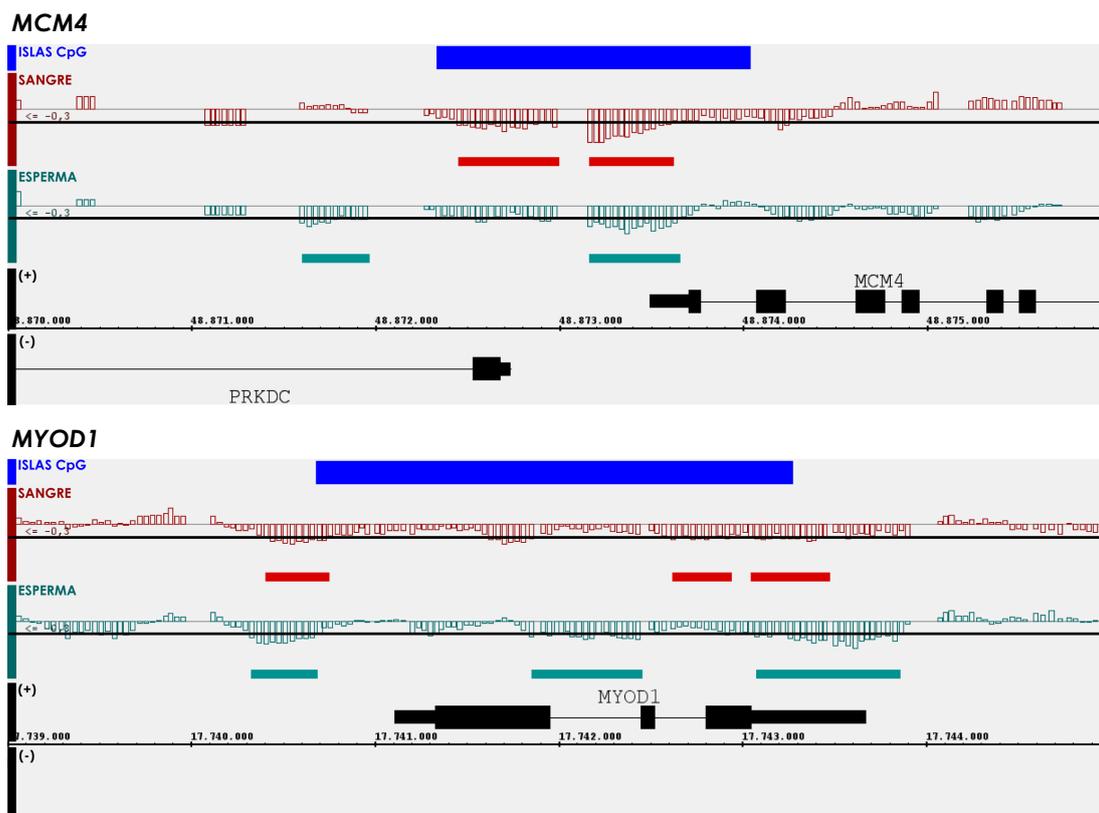


Fig. 35. Análisis comparativo de islas CpG asociadas a genes de expresión constitutiva o genes específicos de tejido. Datos de MeDIP-CHIP para las islas asociadas a los genes *MCM4* y *MYOD1* en DNA de sangre y esperma.

Otro indicativo de la limitación de esta técnica para estimar adecuadamente diferencias pequeñas en los niveles de metilación se observó al analizar dos genes de *imprinting*, *H19* y *KCNQ1*. *H19* se expresa desde el alelo materno y posee una región diferencialmente metilada (DMR) 2 Kb corriente arriba de su promotor, la cual se encuentra metilada en el alelo paterno y no metilada en el materno. *KCNQ1* también se expresa específicamente desde el alelo materno y presenta una isla diferencialmente metilada en uno de sus intrones, desde la cual se inicia la transcripción de un RNA antisentido (*KCNQ1OT1*) en el alelo paterno, donde la isla se

encuentra no metilada. En el alelo materno dicha isla se encuentra metilada, lo que impide la producción del transcrito antisentido y permite la correcta expresión del gen *KCNQ1*. Por tanto, el DMR asociado al gen *H19* debería estar totalmente metilado en esperma y hemimetilado en sangre, mientras que la isla asociada al gen *KCNQ1* tendría que estar desprovista de metilación en esperma y hemimetilada en sangre. Esto implica que en DNA de sangre la señal de ambas islas debería ser muy similar, ya que ambas se encuentran hemimetiladas. Sin embargo, esto no ocurría. Pese a que en ambos casos se observaban diferencias entre la señal de sangre y la de esperma, la intensidad de la señal que correspondía a hemimetilación en el caso de *H19* en sangre era equivalente, o muy similar, a la que correspondía a ausencia total de metilación de *KCNQ1* en esperma (Fig. 36).

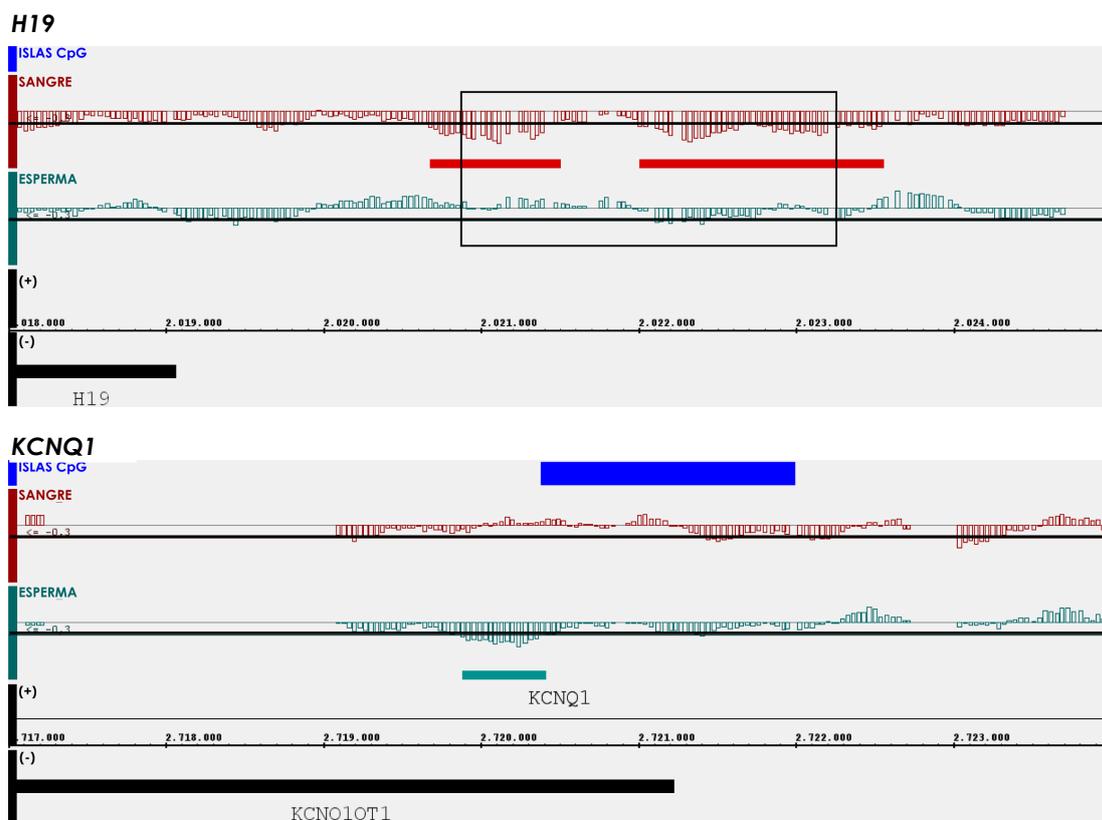


Fig. 36. Análisis comparativo de islas CpG asociadas a genes de imprinting. Datos de MeDIP-CHIP para la isla asociada al gen *KCNQ1* y el DMR (indicado con un recuadro negro) asociado al gen *H19* en DNA de sangre y esperma.

Todas estas observaciones nos han llevado a la conclusión de que la combinación de MeDIP con la hibridación de *tiling microarrays* resulta útil para la detección a nivel genómico de islas CpG metiladas pero no permite estimar de forma fiable niveles bajos o intermedios de metilación, por lo que no es aplicable al objetivo propuesto.

3. DINÁMICA DE GENERACIÓN DE LAS ISLAS CpG EN EL GENOMA DE MAMÍFEROS

3.1 Antecedentes

Desde que se describieron las islas CpG, hace más de veinticinco años (Bird *et al.*, 1985), son muchos los algoritmos que se han propuesto para su identificación a nivel genómico (Tabla 4). Los más antiguos se basan en parámetros asociados a la secuencia de DNA como el tamaño, el contenido en guanina y citosina o la relación entre CpGs observados y esperados (CpG O/E) (Gardiner-Garden & Frommer, 1987; Ponger & Mouchiroud, 2002; Takai & Jones, 2002). Los algoritmos más recientes se basan en parámetros estadísticos para calcular la probabilidad de una secuencia de ser una isla CpG (Hackenberg *et al.*, 2006; Glass *et al.*, 2007). Sin embargo, cuando se han contrastado los datos obtenidos mediante estos algoritmos con datos de metilación derivados de análisis de metilación a nivel genómico (Han & Zhao, 2009; Zhao & Han, 2009), se ha observado que todos presentan falsos positivos, es decir, regiones que no son islas CpG, como ocurre en el caso de muchas repeticiones de la familia Alu. Por otra parte, también se detectan falsos negativos, esto es, regiones que no cumplen los requerimientos para ser clasificadas como islas CpG pero que se encuentran desprovistas de metilación.

ALGORITMO	CRITERIOS
Gardiner-Garden and Frommer (Gardiner-Garden & Frommer, 1987)	Longitud>200pb, G+C>50%, CpG O/E>0,60
Takai and Jones (Takai & Jones, 2002)	Longitud≥500pb, G+C≥55%, CpG O/E≥0,65
CpGPRoD (Ponger & Mouchiroud, 2002)	Longitud>500pb, G+C>50%, CpG O/E>0,60
CpGcluster (Hackenberg <i>et al.</i> , 2006)	Se basa en la distancia entre CpGs contiguos
CG cluster (Glass <i>et al.</i> , 2007)	Humano: ≥27 CpGs en una región ≤ 531 pb Ratón: ≥24 CpGs en una región ≤585 pb

Tabla 4. Principales algoritmos utilizados para la identificación de islas CpG en el genoma de mamíferos.

Esta dificultad a la hora de generar un algoritmo que prediga de forma precisa la presencia de islas CpG en el genoma sugiere que las islas no son regiones que puedan definirse en base a unos parámetros estrictos de tamaño o composición de bases, sino que es posible que en el genoma exista un rango de regiones con un contenido en G+C y CpG O/E, desde el promedio genómico de 40% de G+C y CpG O/E de 0,2 hasta el de las islas CpG canónicas, asociadas con promotores, que satisfacen todos los parámetros establecidos por los diferentes algoritmos. Este grado de degeneración nos indica la posibilidad de que las islas CpG no constituyan una fracción discreta del genoma, y nos ha llevado a cuestionarnos cuáles son los

mecanismos mediante los que se generan las islas CpG en el genoma de los mamíferos. En este sentido, las islas CpG podrían representar el caso extremo de un proceso que eleva el contenido en G+C de determinadas regiones y, eventualmente, previene la metilación de las mismas.

El inicio de la transcripción en las islas podría hacernos pensar que el motivo por el que las islas CpG se han generado en el genoma está vinculado a la regulación de la expresión génica. No obstante, pese a la elevada colocación entre islas CpG y promotores en mamíferos, estas no constituyen elementos esenciales para la regulación de la expresión génica puesto que más del 30% de los genes de mamíferos se transcriben sin necesidad de tener una isla CpG en su promotor. Además, existen muchos organismos, como *Drosophila melanogaster* o *Caenorhabditis elegans*, cuyos genomas carecen de islas CpG.

Numerosas evidencias recopiladas de la literatura, que se comentarán en el apartado de Discusión, sugieren que las islas se han generado como consecuencia de fenómenos de inestabilidad genética asociados, fundamentalmente, al inicio de la transcripción en estas regiones en la línea germinal. Esta inestabilidad se habría traducido en una huella en la secuencia de DNA caracterizada por el incremento en el contenido en guanina y citosina que, una vez alcanzada una determinada densidad de CpGs, haría que estas regiones escaparan a la metilación genómica convirtiéndose en lo que denominamos islas CpG.

Una predicción de este modelo, asumiendo que los mecanismos que han generado las islas CpG siguen operando en la actualidad, es que deberíamos encontrar en el genoma islas en proceso de formación, es decir, regiones que debido a estos procesos de inestabilidad hubieran comenzado a incrementar su contenido en guanina y citosina y que, en aquellos casos en que hubieran alcanzado una densidad de CpGs suficiente, escaparían a la metilación genómica. Con el objetivo de validar esta hipótesis llevamos a cabo varios análisis que se detallan a continuación.

3.2 Identificación y caracterización de regiones no metiladas del genoma que no se clasifican como islas CpG (MeDIP-CHIP)

En primer lugar, haciendo uso de los datos obtenidos mediante la combinación del MeDIP con la hibridación de *tiling microarrays*, decidimos analizar aquellas regiones en las que detectábamos ausencia de metilación pero que no cumplían los criterios mínimos de los algoritmos tradicionales, en cuanto a su composición en G+C o a su frecuencia de dinucleótidos CpG, para clasificarse como islas CpG.

Para seleccionar las regiones a analizar se estableció el mismo umbral utilizado en el apartado 2.5, por el cual se establecieron como regiones no metiladas todas aquellas que presentasen un valor inferior a $-0,3$ para el logaritmo en base 2 del cociente entre el valor del MeDIP sobre DNA de sangre o esperma y el valor del MeDIP sobre DNA metilado *in vitro* ($\log_2(IP/mIP) \leq -0,3$), incrementando en este caso el tamaño mínimo de 400 a 500 pb. Utilizando estos parámetros pudimos comprobar que la mayor parte de las regiones detectadas correspondían con islas CpG. Se contrastaron las regiones obtenidas con las islas CpG predichas por el UCSC Browser, que utiliza el algoritmo de Gardiner-Garden and Frommer (Gardiner-Garden & Frommer, 1987) (Tabla 4), y por la herramienta MapViewer del NCBI (Sayers *et al.*, 2011), que utiliza el algoritmo de Takai and Jones (Takai & Jones, 2002) (Tabla 4), y seleccionamos aquellas regiones que en ninguno de los casos se clasificaban como islas CpG. A continuación se comprobó que las regiones seleccionadas no coincidían con secuencias repetitivas ni se encontraban próximas a estas. Finalmente, se seleccionaron únicamente aquellas regiones que se detectaban como no metiladas tanto en DNA de sangre como en DNA de esperma (Fig.37). Mediante estos criterios de selección encontramos un total de 24 regiones distribuidas en los tres cromosomas analizados (Tabla 5).

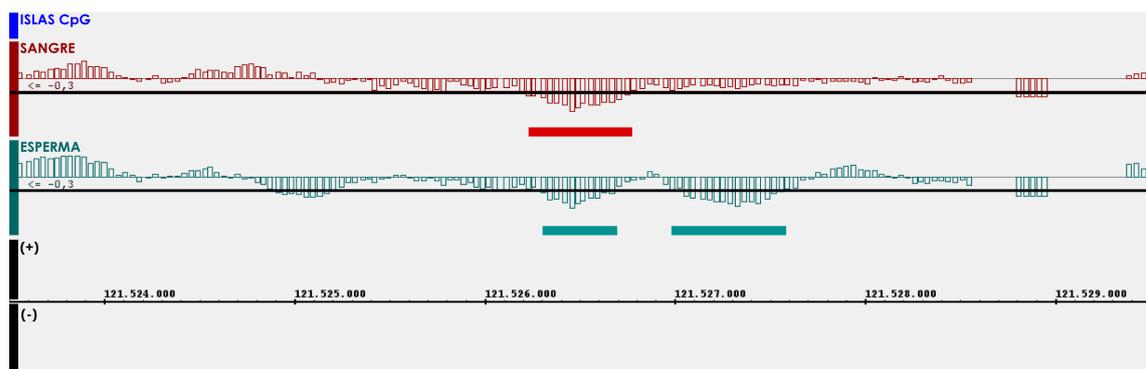


Figura 37. Regiones desprovistas de metilación en sangre y esperma. Datos de MeDIP-CHIP para una región no metilada que no se clasifica como isla CpG. En el DNA de esperma se detecta una segunda región desprovista de metilación que parece encontrarse metilada en el DNA de sangre.

En primer lugar decidimos contrastar nuestros datos de metilación con los obtenidos en los metilomas recientemente publicados (Laurent *et al.*, 2010). Estos metilomas están generados mediante transformación del DNA con bisulfito sódico combinada con secuenciación masiva, lo que ha permitido obtener un mapa de metilación del genoma con resolución de nucleótido. Actualmente se dispone de estos datos para células madre embrionarias (hESC= Human Embryonic Stem Cells), fibroblastos derivados de hESC y fibroblastos de neonato (Fig. 38).

POSICIÓN	hESC	hESC FIBRO	FIBRO	LOCALIZACIÓN	FANTOM	DBTSS	DBTSS FETAL	%G+C	CpG O/E
1 chr8:28,403,595-28,404,247	NM	NM	NM	P	X	X	X	52	0,61
2 chr12:47,545,528-47,546,354	NM	NM	NM	P	X	X	X	56	0,43
3 chr11:28,085,660-28,086,532	NM	NM	NM	P	X	X	X	53	0,60
4 chr12:119,224,324-119,224,922	NM	NM	NM	P	X	X	X	48	0,62
5 chr12:1,6,650,424-1,6,651,044	NM	NM	NM	P	X	X	X	55	0,61
6 chr11:66,764,313-66,765,158	NM	NM	NM	P	X	X	X	51	0,42
7 chr12:1,584,757-1,585,399	NM	NM	NM	I	X	X	X	56	0,49
8 chr11:57,286,308-57,287,049	NM	NM	NM	G	X	X	X	38	0,37
9 chr12:56,142,363-56,143,326	NM	NM	NM	G	X	X	X	55	0,36
10 chr11:66,763,423-66,764,001	NM	NM	NM	G	X			55	0,25
11 chr11:1,121,031,343-1,121,031,890	NM	NM	NM	I				57	0,71
12 chr12:50827046-50827,675	M	NM	NM	I	X			66	0,24
13 chr12:119,247,667-119,248,402	PM	PM	NM	G	X		X	55	0,59
14 chr11:65,391,633-65,392,298	NM	M	M	G	X			56	0,58
15 chr8:1980031-1980972	M	M	NM	P	X	X	X	51	0,47
16 chr12:122,045,024-122,045,849	M	M	M	G			X	61	0,21
17 chr12:121,251,313-121,251,935	M	M	M	G		X	X	63	0,32
18 chr12:112,149,603-112,150,203	M	M	M	G	X			51	0,22
19 chr11:9,42,501-9,43,150	M	M	M	G			X	51	0,32
20 chr11:7,4,733,558-7,4,734,271	M	M	M	G			X	59	0,19
21 chr11:3,4,278,380-3,4,278,950	M	M	M	G	X			41	0,36
22 chr11:118,491,331-118,492,322	M	M	M	G	X	X	X	58	0,51
23 chr8:889,429-890,097	M	M	M	G	X			57	0,57
24 chr8:102,749,013-102,749,723	M	M	M	G	X	X		54	0,22
	12/24	12/24	14/24	7/24	19/24	11/24	14/24	23/24	17/24
	50%	50%	58%	29%	79%	46%	58%	92%	71%

Tabla 5. Análisis de las regiones no metiladas no clasificadas como islas CpG. Se muestran, para cada una de las 24 regiones analizadas, los datos de metilación (NM: no metilado; M: metilado; PM: parcialmente metilado) para tres tipos celulares: hESC, fibroblastos derivados de hESC (hESC-FIBRO) y fibroblastos de neonato (FIBRO). Además se indica la localización de estas regiones en relación con los genes (P: promotor; I: región intergénica; G: región intragénica) y la presencia (X) de sitios de inicio de la transcripción (FANTOM, DBTSS en tejidos somáticos y DBTSS en tejidos fetales) en cada región. Se han sombreado aquellas regiones en las que los parámetros analizados están más próximos a los de las islas CpG.

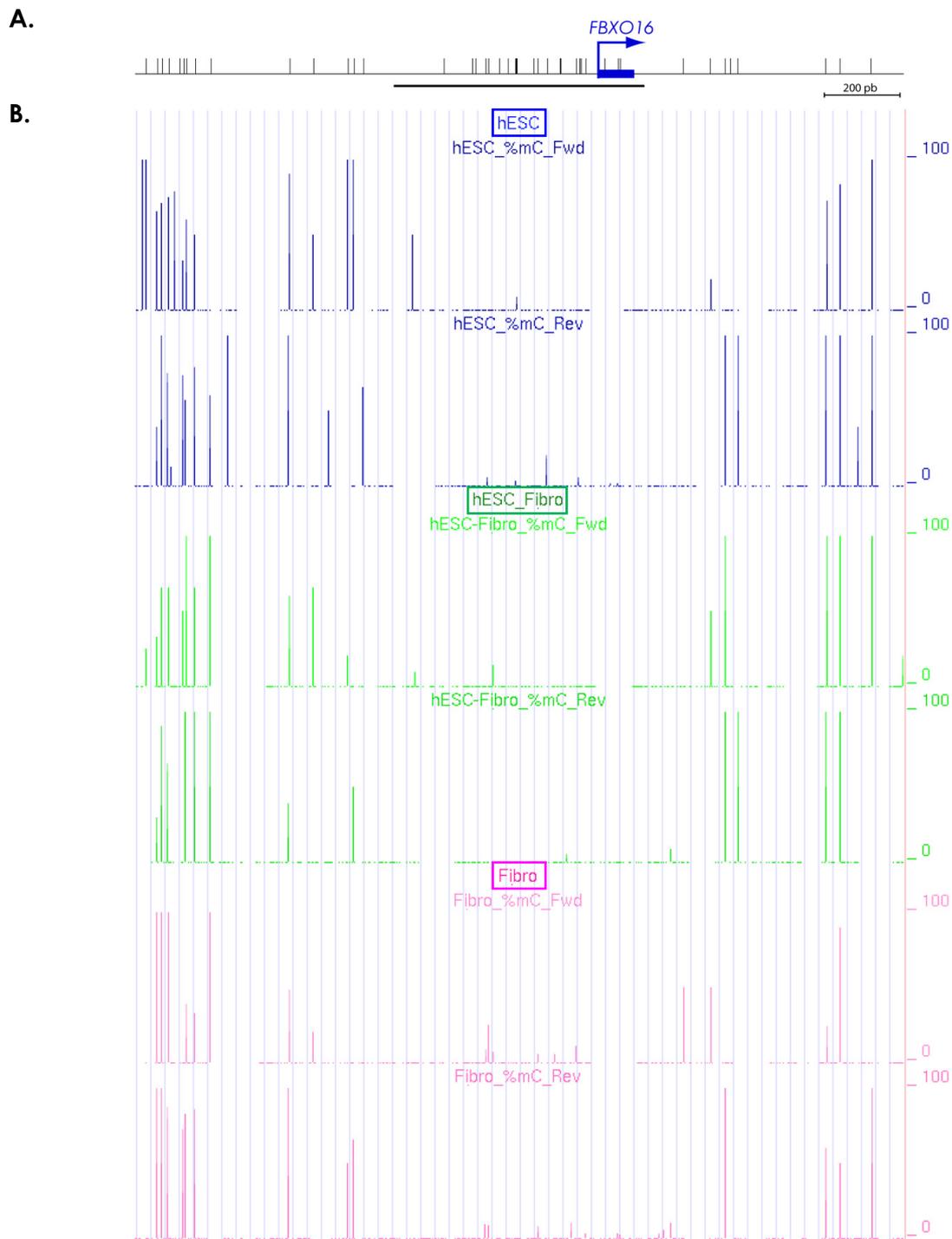


Figura 38. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para una región detectada como desprovista de metilación en sangre y esperma por MeDIP-CHIP. A. Mapa de la región número 1 de la tabla 5, señalada con una línea negra, en el que se representa la posición de los dinucleótidos CpG (líneas verticales). **B.** Patrón de metilación de esta región en las dos cadenas del DNA (cadena líder: Fwd y cadena rezagada: Rev) para los distintos tipos celulares analizados. La escala del eje Y indica el porcentaje de metilación que se representa como una línea vertical para cada uno de los dinucleótidos CpG.

Mediante esta análisis pudimos comprobar que, a pesar de que los tipos celulares son diferentes de los utilizados en nuestro experimento de MeDIP-CHIP, más del 50% de las regiones que identificamos por MeDIP-CHIP se detectaban como no metiladas en alguno de los tres metilomas considerados (**Tabla 5**).

A continuación nos preguntamos si en estas regiones, al igual que las islas CpG, se iniciaba la transcripción. Para comprobar si las regiones seleccionadas contenían promotores de genes anotados utilizamos el navegador UCSC *Browser*, que integra los datos del Gen Bank (Benson *et al.*, 2008). Aproximadamente un 30% de las regiones analizadas correspondían a promotores anotados de genes, mientras que el resto correspondía a regiones intergénicas (12,5%) e intragénicas (57,5%) (**Tabla 5**). Seguidamente, recurrimos a dos bases de datos, DBTSS (Suzuki *et al.*, 2002) y FANTOM (Kawaji *et al.*, 2009), en las que se describen sitios de inicio de la transcripción localizados utilizando diferentes aproximaciones (Ver materiales y métodos) (**Fig. 39**). Como puede observarse en la tabla 5, veintitrés de las veinticuatro regiones presentaban sitios de inicio de la transcripción en alguna de las dos bases de datos consultadas.

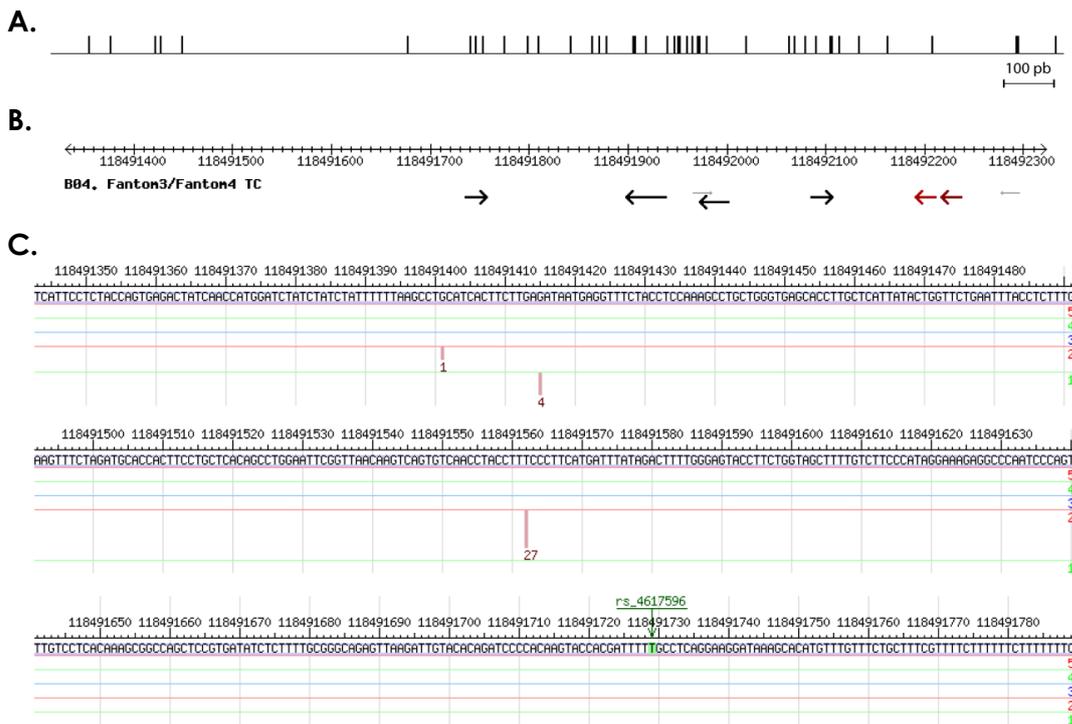


Figura 39. Mapa de sitios de inicio de la transcripción (TSS) en una de las regiones desprovistas de metilación en sangre y esperma que no se clasifica como isla CpG. A. Mapa de la región 22 de la tabla 5 en el que se representa la posición de los dinucleótidos CpG. **B.** Mapa de TSSs de la región 22 según FANTOM. Cada flecha indica la presencia de un sitio de inicio de la transcripción y la orientación de la misma. El color varía entre rojo intenso y negro en función de que el nivel de expresión sea alto o bajo, respectivamente. **C.** Mapa de TSSs de la región 22 según DBTSS. Se señalan los TSSs indicando en cada caso el número de clones detectado. Las líneas horizontales enumeradas del 1 al 5 a la derecha de cada secuencia representan cada uno de los tejidos analizados.

Finalmente, analizamos el contenido en G+C de las 24 regiones identificadas, así como la proporción CpG O/E. Encontramos que más del 90% de estas regiones poseía un contenido en guanina y citosina superior al contenido promedio del genoma, del 40%. Además, en torno a un 70% de ellos presentaba un CpG O/E que se encuentra muy por encima del 0,2 que constituye el promedio genómico (**Tabla 5**).

En conjunto, estos datos sugerían que las regiones analizadas presentaban características que se encontraban más próximas a las islas CpG que al resto del genoma de modo que, de acuerdo con nuestro modelo, podría tratarse de islas que se están generando en el genoma humano.

3.3 Identificación y análisis de una isla CpG en proceso de formación en el genoma de *Mus musculus*

El locus pseudoautosómico de los cromosomas X e Y de mamíferos es una región de homología a través de la cual ambos cromosomas se aparean y que, por tanto, está sometido a elevadas tasas de recombinación meiótica, siendo una zona de gran inestabilidad.

El gen *FXY*, también denominado *MID1*, es un gen ligado al cromosoma X en humano que codifica una proteína de 667 aminoácidos, *MID1*, muy conservada en mamíferos, a excepción de la especie de ratón *Mus musculus* (Perry & Ashworth, 1999). *MID1* posee actividad ubiquitin-ligasa y se ha demostrado que se asocia a los microtúbulos, contribuyendo a la estabilización de los mismos (Schweiger *et al.*, 1999; Aranda-Orgillés *et al.*, 2008). Además su mutación se ha relacionado con el síndrome de Opitz, que se caracteriza por diversas malformaciones que en la mayor parte de los casos van acompañadas de retraso mental (Quaderi *et al.*, 1997).

En *Mus musculus* este gen sufrió una traslocación por la que el extremo 3' quedó situado dentro del locus pseudoautosómico mientras que la región 5' permaneció en la parte específica del cromosoma X (Perry & Ashworth, 1999). Esta traslocación no ha tenido lugar en otras especies de mamíferos como humano o rata ni tampoco en la especie de ratón *Mus spretus* (**Fig. 40**) que divergió de *Mus musculus* hace unos tres millones de años, lo que indica que dicha traslocación tuvo lugar en un momento relativamente reciente en la evolución. Se trata, por tanto, de una región sometida a gran inestabilidad en el genoma de *Mus musculus* en la que además se ha detectado un incremento en el contenido en guanina y citosina con respecto a su ortóloga en otros mamíferos (Perry & Ashworth, 1999; Montoya-Burgos *et al.*, 2003). Esto convierte a esta región en un buen modelo para el estudio de la hipótesis que hemos

propuesto por la que la inestabilidad genética en el entorno de las islas sería la responsable del incremento en la composición de G+C que caracteriza a estas regiones.

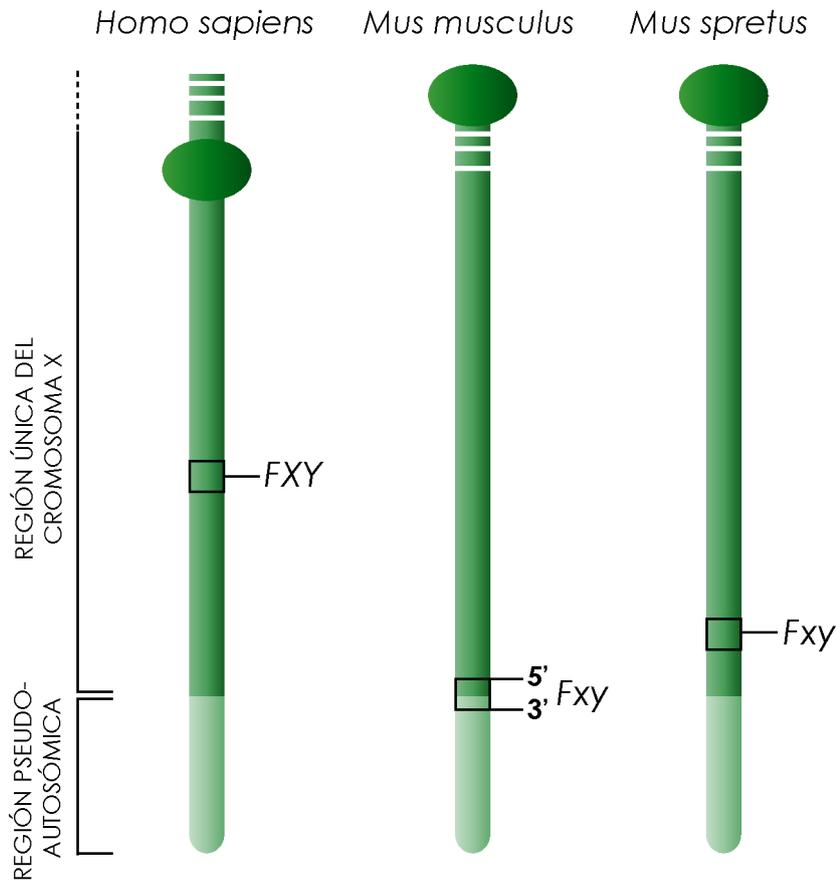


Figura 40. Posición del gen *Fxy* en el cromosoma X de *Homo sapiens*, *Mus musculus* y *Mus spretus*. Se muestra la localización del gen *Fxy* en el cromosoma X de cada una de las tres especies. Se señala además la posición de la región pseudoautosómica y la región única del cromosoma X. Las elipses representan la posición del centrómero en cada caso.

Mediante un análisis detallado del gen *Fxy* en ratón (*Mus musculus*), humano y rata pudimos comprobar cómo la región 5' del gen, que se encuentra en todas las especies en la región específica del cromosoma X, mostraba valores, tanto en el contenido en guanina y citosina como en la proporción CpG O/E, que se situaban en torno al promedio genómico. Sin embargo, la región 3' del gen mostraba un notable incremento en estos parámetros en el ratón (*Mus musculus*) (**Fig. 41**) pero mantenía estos valores similares a la región 5' en humano y rata.

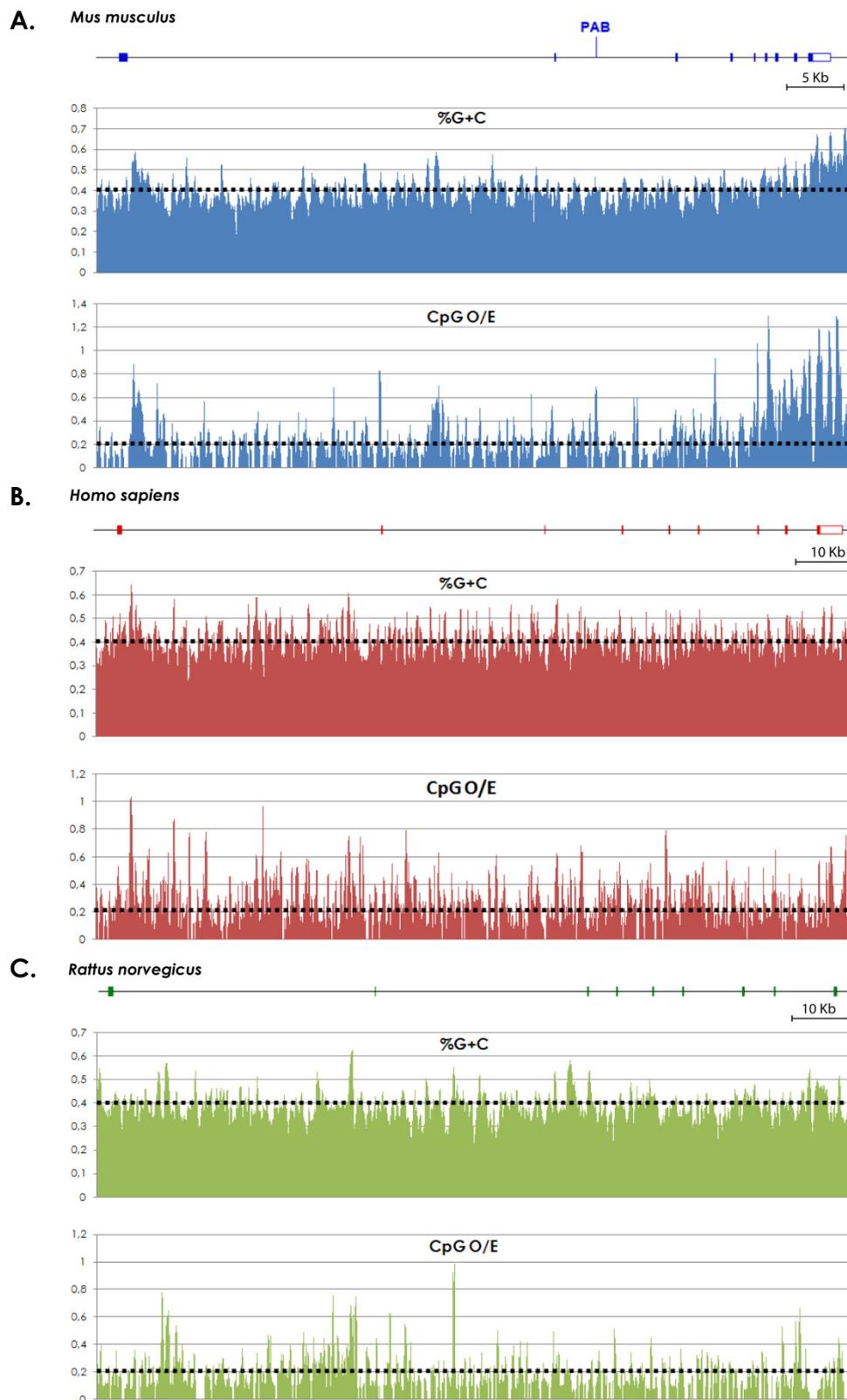


Figura 41. Contenido en G+C y ratio CpG O/E del gen *Fxy*. Se representa el mapa del gen *Fxy* para cada especie: *Mus musculus* (A), *Homo sapiens* (B) y *Rattus norvegicus* (C). En el mapa correspondiente a *M. musculus* se señala, además, el límite de la región pseudoautosómica (PAB: *Pseudoautosomal Boundary*). La línea de puntos señala el valor promedio del genoma para el porcentaje de G+C y para la relación CpG O/E.

Ese incremento en el contenido en guanina y citosina en *M. musculus* se hizo patente al realizar un alineamiento múltiple entre las secuencias ortólogas de varios mamíferos para los exones 1 y 9 del gen *Fxy*. Al analizar el exón 1, situado en la región específica del cromosoma X, pudimos observar cómo el número de sustituciones detectadas en otras especies con respecto a la secuencia de *M. musculus* aumentaba proporcionalmente a la distancia evolutiva. Por ejemplo, se observaba un total de 40 sustituciones en humano, la especie filogenéticamente más alejada, mientras que sólo había 3 sustituciones en *M. spretus*, que pertenece al mismo género que *M. musculus* (Tabla 6). El exón 9, por su parte, mostraba un número de sustituciones mucho mayor que el exón 1 en *Mus musculus* con respecto al resto de especies consideradas. En este caso, incluso se observaba una mayor conservación entre el humano y la rata que entre las dos especies de ratón, a pesar de pertenecer al mismo género (Tabla 6). Mediante este análisis pudimos comprobar, además, que la inmensa mayoría de las sustituciones que tenían lugar en el exón 9 de *M. musculus* iban en la dirección A/T → C/G, es decir, se producía una ganancia en G+C con respecto a la región ortóloga de los otros mamíferos analizados (Tabla 6). Esto no ocurría en el caso del exón 1, donde las sustituciones A/T → C/G y C/G → A/T tenían lugar en proporciones similares.

SUSTITUCIONES	A/T	G/C	G/C	A/T
	↓ G/C	↓ A/T	↓ C/G	↓ T/A
EXÓN 1				
<i>M. spretus</i> → <i>M. musculus</i>	1	1	1	0
<i>R. norvegicus</i> → <i>M. musculus</i>	12	6	0	0
<i>S. vulgaris</i> → <i>M. musculus</i>	10	19	2	2
<i>H. sapiens</i> → <i>M. musculus</i>	17	20	0	3
<i>R. norvegicus</i> → <i>H. sapiens</i>	14	26	1	2
EXÓN 9				
<i>M. spretus</i> → <i>M. musculus</i>	59	0	2	0
<i>R. norvegicus</i> → <i>M. musculus</i>	63	3	2	0
<i>S. vulgaris</i> → <i>M. musculus</i>	59	6	4	0
<i>H. sapiens</i> → <i>M. musculus</i>	51	7	7	0
<i>R. norvegicus</i> → <i>H. sapiens</i>	9	23	2	2

Tabla 6. Análisis del patrón de sustituciones de los exones 1 y 9 del gen *Fxy* del ratón de laboratorio (*Mus musculus*) con respecto a sus regiones ortólogas en ratón de campo (*Mus spretus*), rata (*Rattus norvegicus*), ardilla (*Sciurus vulgaris*) y humano (*Homo sapiens*). En el exón 9 se aprecia un claro sesgo hacia el incremento en G+C (Sustituciones de tipo A/T → G/C, sombreadas) en *M. musculus* con respecto al resto de especies analizadas.

La elevada relación de CpG O/E en el extremo 3' del gen *Fxy* en *M. musculus* con respecto al promedio genómico (Fig. 41A) indicaba que, o bien esta región escapa a la metilación genómica o bien la frecuencia con la que se generan CpGs debido a la inestabilidad en este locus es mayor a la frecuencia con la que se pierden por desaminación espontánea debida a la metilación. Otra evidencia que sugería que esta región podría escapar a la metilación era la presencia de múltiples sitios de inicio de la transcripción, pese a tratarse del extremo 3' del gen (Fig. 42A). Sin embargo, al analizar con el mismo criterio la

región ortóloga en humano se apreciaba que la iniciación de la transcripción era prácticamente indetectable (**Fig. 42B**).

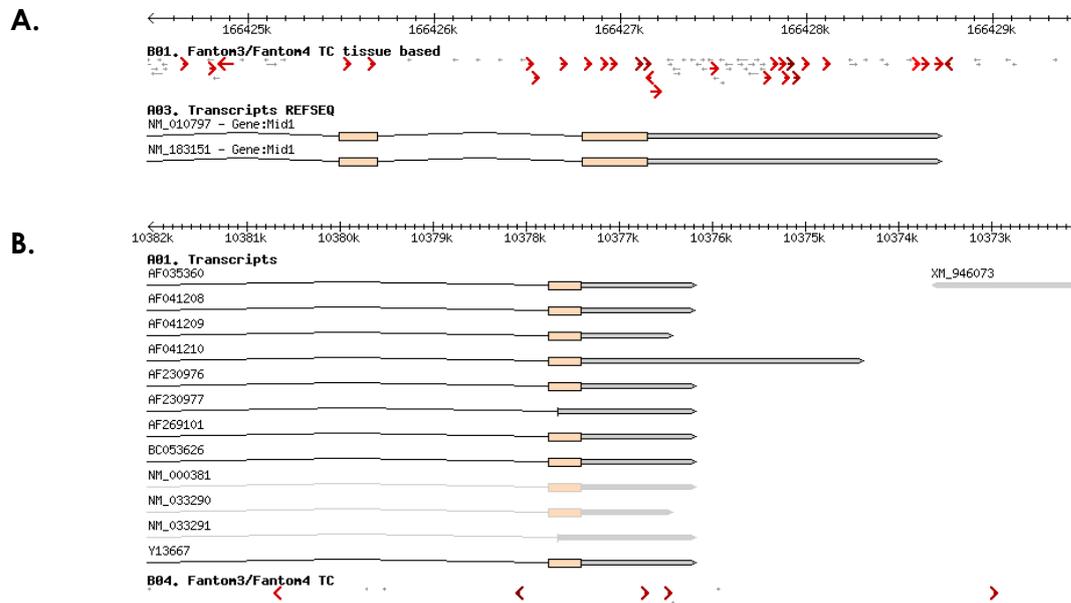


Figura 42. Mapa de sitios de inicio de transcripción en la región 3' del gen *Fxy* en *M. musculus* (A) y *H.sapiens* (B).

Con el objetivo de determinar el estado de metilación de esta región en *M. musculus* llevamos a cabo un análisis de metilación por bisulfito sódico. Para ello utilizamos DNA procedente de 2 muestras de cola y de 5 muestras de células madre embrionarias de ratón (células ES). Se analizó el DNA transformado por bisulfito mediante PCR con oligonucleótidos específicos para una región de 435 pb, correspondiente al último exón del gen *Fxy*, que contiene 42 CpGs. A continuación se clonaron los productos de PCR en *E.coli* y se secuenciaron varios clones procedentes de cada muestra de DNA. Como puede observarse en la figura 43B, el DNA procedente de las colas de ratón presentaba un porcentaje de metilación próximo al 100%. Sin embargo, en el caso de las células ES pudimos observar que en las 5 muestras independientes que analizamos detectábamos clones con niveles variables de metilación, desde clones prácticamente desprovistos de metilación hasta clones totalmente metilados (**Fig. 43A**). Al realizar un cómputo global de todas las secuencias de células ES analizadas pudimos comprobar que el porcentaje de metilación promedio de esta región era del 58%. Un porcentaje similar se observó al analizar de forma independiente el porcentaje de metilación promedio para cada CpG (**Fig. 43C**).

Estos resultados sugieren que existe cierta protección frente a la metilación de esta región en células indiferenciadas como las células ES, mientras que en células diferenciadas como las que forman parte de la cola de ratón el mecanismo responsable de esa protección no parece estar operativo.

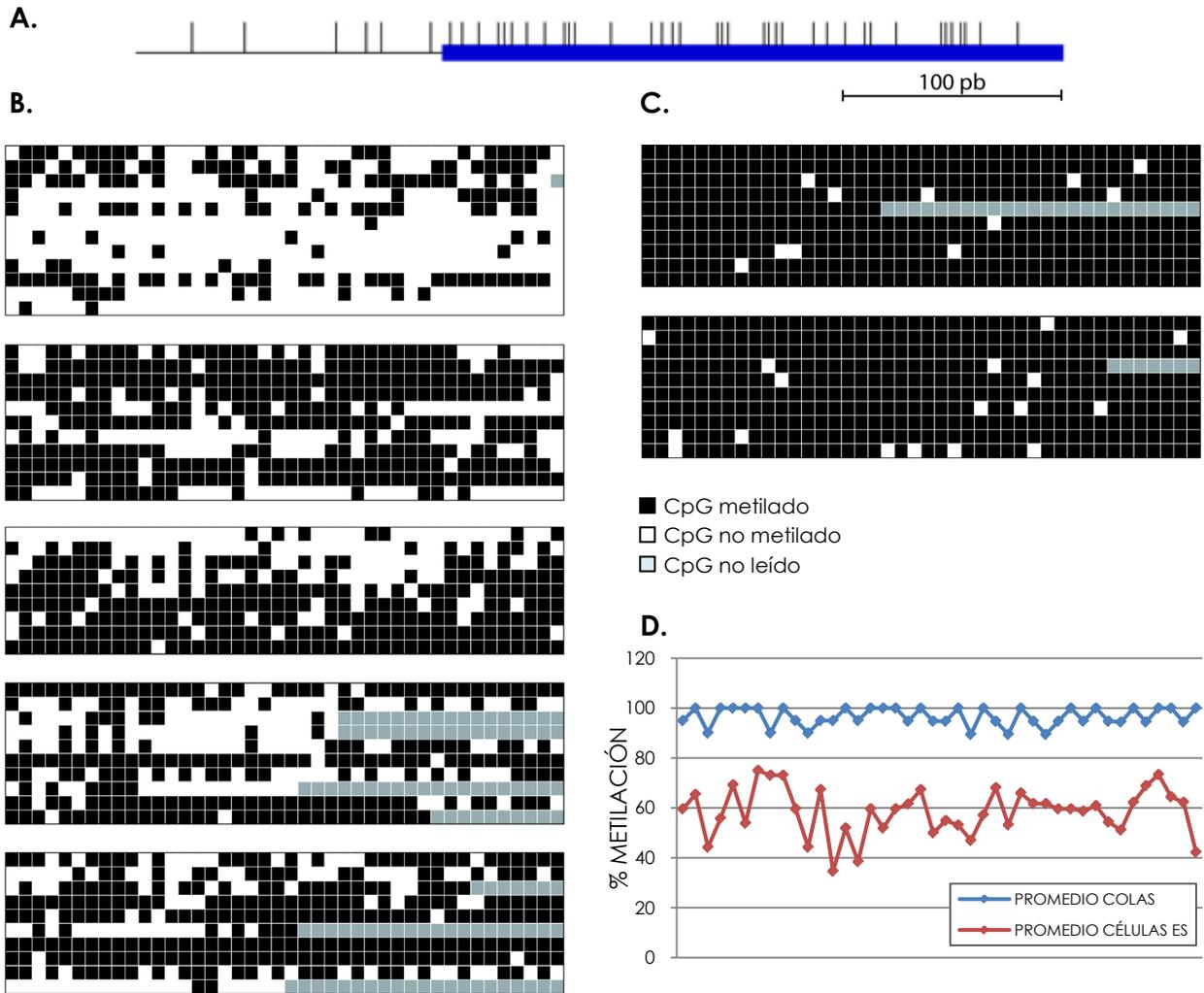


Figura 43. Análisis de metilación del exón 9 del gen *Fxy* de ratón (*M. musculus*). **A.** Mapa de la región analizada por bisulfito en el que se representa la posición de los dinucleótidos CpG (líneas verticales). **B.** Patrón de metilación de los distintos clones analizados para cada una de las cinco muestras de DNA de células ES. Cada línea horizontal corresponde a un clon. **C.** Patrón de metilación de los distintos clones analizados para las dos muestras de DNA procedentes de cola de ratón. **D.** Porcentaje de metilación promedio para cada uno de los 42 CpGs de la región analizada en el total de clones analizados para DNA de células ES y DNA de colas de ratón.

Finalmente, quisimos conocer el estado de metilación del extremo 3' del gen *Fxy* en otras especies, humano y *M. spretus*, en las que no se hubiera producido su traslocación al locus pseudoautosómico y, por tanto, no hubieran sufrido un incremento en el contenido en guanina y citosina en esta región. Para ello recurrimos a los datos de los metilomas humanos para hESC, fibroblastos derivados de hESC y fibroblastos de neonato. Como puede observarse en la figura 44 en los tres casos la región correspondiente al exón 9 del gen *FXY* se encuentra densamente metilada, como cabía esperar dado que en humanos la región 3' del gen *FXY* presenta valores de contenido en G+C y CpG O/E similares a los del promedio genómico (Fig. 41B).

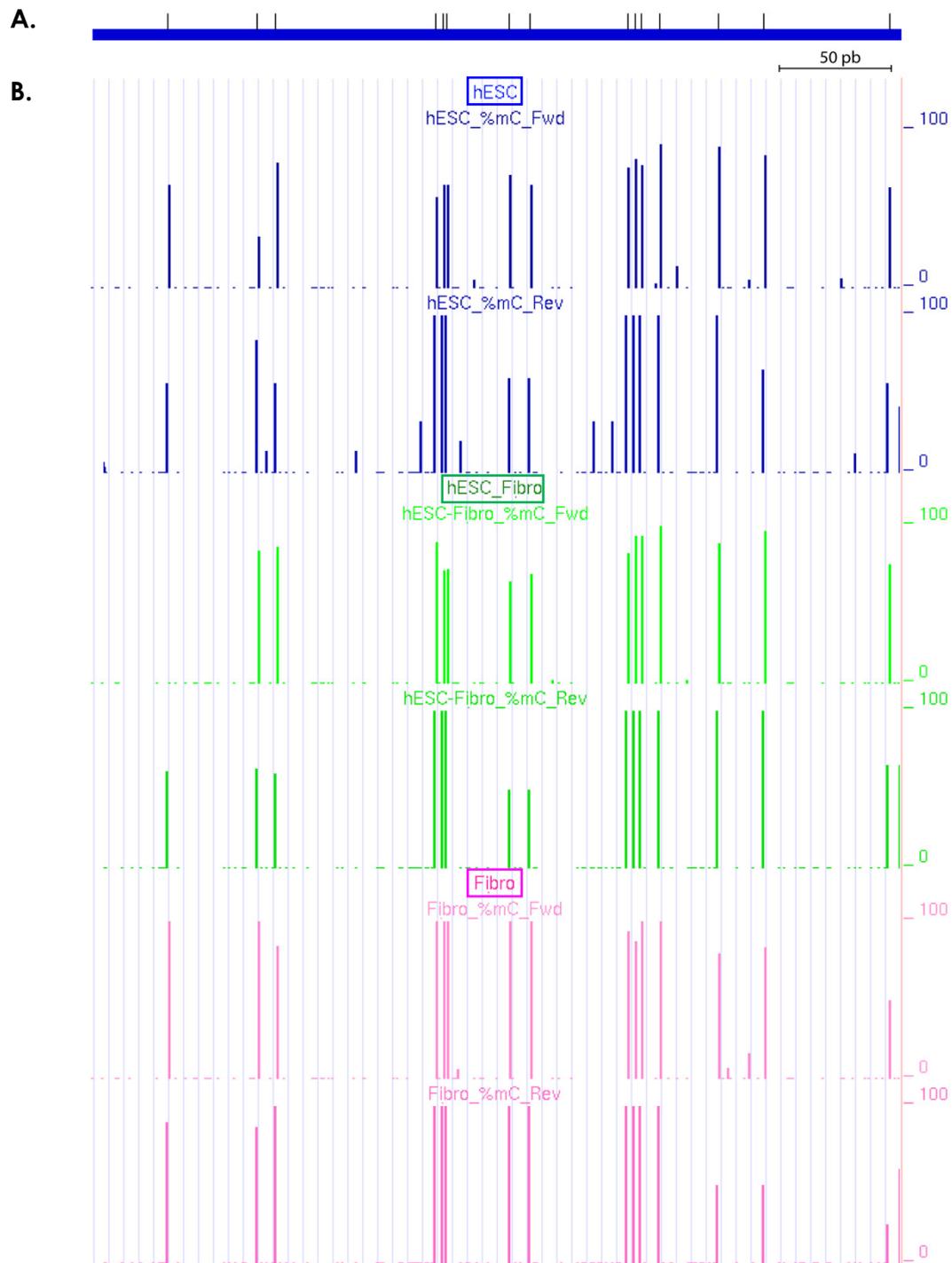
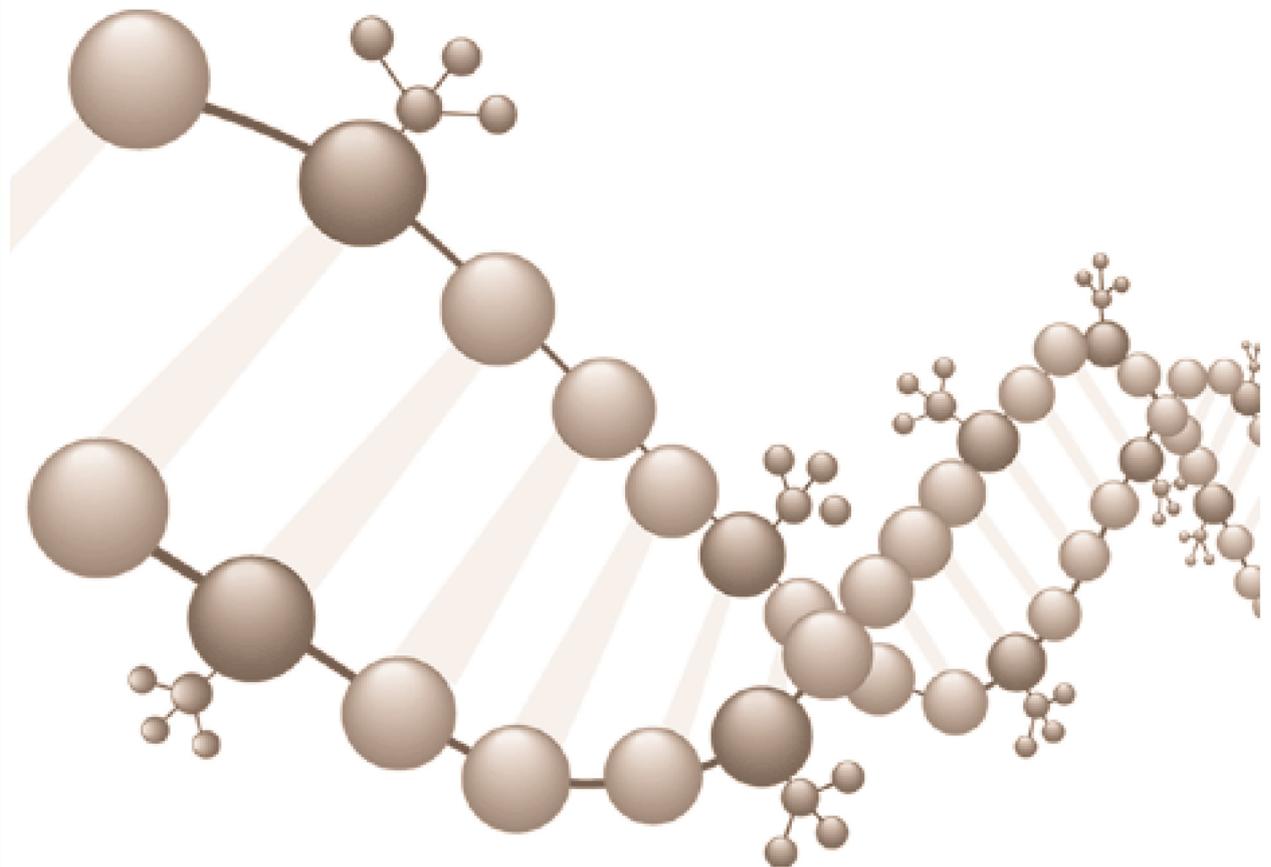


Figura 44. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para el exón 9 del gen FXY humano. A. Mapa de la región del exón 9 del gen FXY humano en el que se representa la posición de los dinucleótidos CpG (líneas verticales). **B.** Patrón de metilación de esta región en las dos cadenas del DNA (cadena líder: Fwd y cadena rezagada: Rev) para los distintos tipos celulares analizados. La escala del eje Y indica el porcentaje de metilación que se representa como una línea vertical para cada uno de los dinucleótidos CpG.

Por otra parte, se llevó a cabo el análisis de metilación por bisulfito sódico del extremo 3' del gen *Fxy* de *Mus spretus* sobre dos muestras de DNA de cola. Se analizó el DNA transformado por bisulfito mediante PCR con oligonucleótidos específicos para una región de 305 pb, correspondiente al último exón del gen *Fxy*, que contiene 8 CpGs. A continuación se secuenció el producto de PCR y pudimos comprobar que todos los CpGs de la región analizada se encontraban metilados (datos no mostrados).

En conjunto, estos datos indican una correlación entre el incremento en el contenido en guanina y citosina como consecuencia de la inestabilidad generada en el extremo 3' del gen *Fxy* en el genoma de *M. musculus* y la protección de esta región frente a la metilación, al menos en células ES. Esta protección no tiene lugar en otras especies en las que no ha tenido lugar este cambio en la composición de bases. No obstante, también se pone de manifiesto que la riqueza en guanina y citosina no es suficiente para proteger a una región frente a la metilación en tejidos diferenciados, puesto que la misma región que escapa parcialmente a la metilación en células ES se encuentra densamente metilada en el DNA de cola. La metilación somática de esta región, sin embargo, no tendrá consecuencias en su composición de bases a nivel evolutivo, puesto que las mutaciones que pudieran producirse en tejidos somáticos por desaminación de citosinas metiladas no pasarán a la siguiente generación.

DISCUSIÓN



1. PAPEL DE LA REPLICACIÓN DEL DNA EN LA PROTECCIÓN FRENTE A LA METILACIÓN DE LAS ISLAS CpG

Cuando se inició este trabajo existían varias evidencias en la literatura que sugerían que la replicación se iniciaba frecuentemente en regiones que contenían islas CpG (Delgado *et al.*, 1998; Keller *et al.*, 2002; Ladenburger *et al.*, 2002). Esta alta colocación entre islas CpG y ORIs se ha puesto de manifiesto posteriormente en varios mapeos de ORIs realizados mediante la combinación del análisis de intermediarios de replicación en cadenas nacientes de DNA con la hibridación de *microarrays* (Cadoret *et al.*, 2008; Sequeira-Mendes *et al.*, 2009). En estos trabajos se ha observado que entre un 40 y un 50% de los ORIs descritos colocan con islas CpG, un porcentaje muy superior al que se esperaría por azar, teniendo en cuenta que, en su conjunto, las islas CpG representan apenas un 1% del genoma.

La elevada conservación filogenética detectada en los ORIs *CDKN2B* y *LMNB2* humanos (**Fig. 6 y 7** Pág. 34 y 35) nos llevó a preguntarnos si ésta sería una característica habitual de los ORIs de mamíferos, o al menos de los ORIs asociados a islas CpG. Por ello, analizamos el grado de conservación filogenética en otras siete regiones asociadas a islas CpG en las que se había mapeado el inicio de la replicación. El resultado de este análisis mostró que en seis de las nueve regiones analizadas, incluyendo los ORIs *CDKN2B* y *LMNB2*, se detectaba un alto grado de conservación en mamíferos a pesar de tratarse de regiones no codificantes (**Fig. 8** Pág. 37 y **Fig. S1 A-F** Pág. 145-150). Este resultado nos hizo preguntarnos si seríamos capaces de detectar ORIs en el genoma mediante la búsqueda de regiones conservadas no codificantes asociadas a islas CpG. Para tratar de responder a esta cuestión analizamos 21 Mb del genoma en búsqueda de regiones conservadas no codificantes en el extremo 5' de islas CpG. Mediante este análisis pudimos comprobar que más del 40% de las islas analizadas muestran una región conservada en su extremo 5' (**Fig. 9** Pág. 38). A continuación intentamos determinar si tres de estas regiones, seleccionadas al azar, unían la maquinaria de inicio de replicación. Para ello llevamos a cabo la inmunoprecipitación de cromatina con anticuerpos dirigidos contra ORC2 y CDC6, dos proteínas que forman parte de los complejos pre-replicativos. Los resultados obtenidos indicaban que, al menos en dos de las tres islas analizadas, la región conservada no coincidía con la zona que presenta una mayor unión de ORC2 y CDC6 (**Fig.15** Pág. 45 y 46). No obstante, la baja eficiencia obtenida en la inmunoprecipitación de cromatina no nos permite asegurar que estos resultados sean significativos. Por ello, basándonos únicamente en estos datos, no podemos afirmar ni descartar que la conservación filogenética en el entorno de las islas CpG pueda servir como indicador de la presencia de un ORI.

Por otra parte, varias evidencias recopiladas de la literatura así como observaciones realizadas en nuestro laboratorio sugerían una posible implicación del inicio de la replicación en la protección de las islas CpG frente a la metilación. Pese a que el número de ORIs que se habían descrito cuando se inició este trabajo era escaso, existían varios indicios en la literatura que sugerían que la replicación se iniciaba frecuentemente en regiones que contenían islas CpG (Delgado *et al.*, 1998; Keller *et al.*, 2002; Ladenburger *et al.*, 2002). Además parecía existir cierta tendencia a que los ORIs que colocalizaban con islas se situaran en el extremo 5' de las mismas (Giacca *et al.*, 1994; Taira *et al.*, 1994; Keller *et al.*, 2002; Ladenburger *et al.*, 2002; Ghosh *et al.*, 2006; Gonzalez *et al.*, 2006). Asimismo, se ha demostrado que algunas subunidades de ORC permanecen unidas a los ORIs durante todo el ciclo celular (Kreitz *et al.*, 2001; Abdurashidova *et al.*, 2003). Esto, unido a la observación de que las islas CpG son más proclives a metilarse por su extremo 3' (Ghazi *et al.*, 1992; Rideout *et al.*, 1994) (Cuadrado, Delgado & Antequera, sin publicar), nos llevó a plantearnos la posibilidad de que la unión preferente de ORC en el extremo 5' de las islas CpG podría contribuir a la protección de estas frente a la metilación. Esta hipótesis se vio reforzada tras analizar el estado de metilación de las islas CpG asociadas a los ORIs *CDKN2B* y *LMNB2* y comprobar que el límite de metilación de estas islas en 5' coincidía exactamente con la región en la que se había descrito la unión de ORC. Estos datos sugerían que ORC podía estar actuando como una barrera frente a la metilación en el extremo 5' de las islas CpG.

A continuación quisimos saber si las regiones que habíamos seleccionado como ORIs potenciales en base a su conservación filogenética (Apartado 1.2 de Resultados) también establecían el límite de metilación de las islas CpG asociadas. Para ello seleccionamos al azar cinco de estas regiones y llevamos a cabo su análisis de metilación por bisulfito sódico. Como se observa en la figura 13, en todos los casos la región conservada marcaba el límite de metilación en el extremo 5' de las islas. Al contrastar estos resultados con los datos de unión de ORC2 y CDC6 obtenidos por ChIP para tres de estas regiones, pudimos comprobar que, al menos en dos de ellas, la región conservada no coincidía con la de mayor enriquecimiento para estas proteínas (**Fig. 15** Pág. 45 y 46), en contra de lo que ocurría para los ORIs *CDKN2B* y *LMNB2*. En una de estas regiones se comprobó mediante el análisis de intermediarios de replicación en cadenas nacientes de DNA que el inicio de la replicación coincidía con la región en la que se detectaba el máximo enriquecimiento para ORC2 y CDC6 por ChIP (**Fig. 17** Pág. 48).

Varios trabajos publicados con posterioridad a nuestro análisis ratificaron que es frecuente detectar un alto grado de conservación filogenética en la secuencia de las regiones en las que se ha descrito el inicio de la replicación en humanos. El mapeo de ORIs a lo largo

del 1% del genoma permitió a Cadoret *et al.* (2008) y a Karnani *et al.* (2010) la identificación de 283 y 150 nuevos ORIs en el genoma humano, respectivamente. Un 70% de los ORIs identificados por Cadoret *et al.* (2008) y más del 50% de los que mapean Karnani *et al.* (2010) incluyen regiones evolutivamente conservadas identificadas en el genoma de mamíferos por el proyecto ENCODE (Margulies *et al.*, 2007). Esta conservación filogenética es mayor en los ORIs asociados a sitios de inicio de la transcripción, sobre todo aquellos que coinciden con islas CpG. Pero incluso en los ORIs no asociados a TSSs la conservación que se observa es mayor de la que cabría esperar al azar (Necsulea *et al.*, 2009). El hecho de que la secuencia de las regiones que funcionan como ORIs en humano se encuentre frecuentemente conservada en otros mamíferos sugiere que podría existir también una conservación funcional, es decir, que la localización de los ORIs en mamíferos podría estar conservada, al menos parcialmente, y por tanto también podría estarlo el mecanismo mediante el cual se especifican los sitios en los que se inicia la replicación. De acuerdo con esta hipótesis se ha observado que el patrón temporal de iniciación de la replicación está bastante conservado en regiones sinténicas entre humano y ratón (Farkash-Amar *et al.*, 2008; Ryba *et al.*, 2010).

La ausencia de una secuencia consenso entre los ORIs caracterizados hasta la fecha, unida al hecho de que ORC no muestra especificidad de secuencia (Vashee *et al.*, 2003; Schaarschmidt *et al.*, 2004) indican que la frecuente conservación filogenética observada en los ORIs no parece deberse a que existan requerimientos específicos de secuencia para la unión de ORC. Sin embargo, esta conservación podría ser indicativa de la unión específica de otras proteínas que tal vez contribuirían al reclutamiento indirecto de ORC a estas regiones del genoma.

Sería interesante determinar qué proteínas se unen a estas regiones conservadas de cara a arrojar algo de luz sobre el modo en que se especifican los ORIs en el genoma de mamíferos. Cadoret *et al.* (2008) han observado que existe un enriquecimiento en sitios de unión para los factores de transcripción C-Jun y C-Fos en los 283 ORIs que describen en su estudio (Cadoret *et al.*, 2008). Además, existen evidencias de la interacción entre factores de transcripción y ORC en distintos organismos (Bosco *et al.*, 2001; Beall *et al.*, 2002; Minami *et al.*, 2006). Por otra parte, la asociación preferente de los ORIs con sitios de inicio de la transcripción y una estructura accesible de la cromatina (enriquecida en marcas activadoras de histonas y sitios de hipersensibilidad a nucleasas) sugiere que la especificación de las regiones en las que se inicia la replicación en mamíferos está muy influenciada por el inicio de la transcripción y la estructura de la cromatina asociada a esta, de modo que ORC podría aprovechar estos entornos abiertos y accesibles que se generan en los promotores para unirse al DNA e iniciar la

replicación (Audit *et al.*, 2009; Necsulea *et al.*, 2009; Sequeira-Mendes *et al.*, 2009; Karnani *et al.*, 2010).

Tomados en conjunto, nuestros datos indican que, pese a las observaciones realizadas en los ORIs *CDKN2B* y *LMNB2* la unión de ORC en el extremo de las islas CpG no parece ser el principal mecanismo responsable de la protección frente a la metilación de las mismas. Por otra parte, si el inicio de la replicación tuviera un papel relevante en la protección frente a la metilación todos los ORIs deberían encontrarse desprovistos de metilación, no sólo en aquellos asociados a islas CpG. Sería interesante caracterizar el estado de metilación de todos los ORIs descritos haciendo uso de los metilomas de los que se dispone actualmente (Lister *et al.*, 2009; Laurent *et al.*, 2010), para tratar de esclarecer cuál es la contribución real del inicio de la replicación a la protección de las islas CpG frente a la metilación. No obstante, se ha observado que en las islas CpG situadas en el cromosoma X inactivo en hembras de ratón, que se encuentran metiladas y transcripcionalmente inactivas, el inicio de la replicación es comparable al del cromosoma X activo (Cohen *et al.*, 2003), salvo porque se produce un poco más tarde durante la fase S (Gomez & Brockdorff, 2004). Esta evidencia indica que la metilación no previene la unión de ORC o la actividad de los ORIs, y apunta a que el inicio de la replicación probablemente no es el principal mecanismo responsable del estado no metilado de las islas CpG.

2. CONTRIBUCIÓN DE LA TRANSCRIPCIÓN AL MANTENIMIENTO DE LAS ISLAS CpG EN ESTADO NO METILADO

2.1 Metilación diferencial de las islas CpG y regulación de la expresión génica

La metilación de una isla CpG asociada a un promotor conlleva el silenciamiento del mismo (Eckhardt *et al.*, 2006; Weber *et al.*, 2007). No obstante, la mayor parte de las islas CpG permanecen libres de metilación incluso en aquellos tejidos en los que el gen asociado no se está expresando (Weber *et al.*, 2007; Rakyan *et al.*, 2008; Suzuki & Bird, 2008). Esto sugiere que la metilación del DNA, aunque tiene una contribución indiscutible en la represión transcripcional, no parece constituir el principal mecanismo para el silenciamiento génico en la célula. De hecho, numerosas evidencias señalan que la metilación constituye un mecanismo de silenciamiento estable de regiones previamente silenciadas por otros medios (Wutz & Jaenisch, 2000; Feldman *et al.*, 2006).

El reciente desarrollo de técnicas de análisis de metilación a nivel genómico ha puesto de manifiesto que, además de las islas CpG localizadas en el cromosoma X inactivo de hembras y las que se asocian a genes de *imprinting*, existe un porcentaje de islas CpG que se encuentran metiladas en cada tejido (Yamada *et al.*, 2004; Weber *et al.*, 2005; Eckhardt *et al.*, 2006; Weber *et al.*, 2007; Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Rauch *et al.*, 2009; Straussman *et al.*, 2009; Illingworth *et al.*, 2010; Maunakea *et al.*, 2010). Un análisis detallado del patrón de expresión de los genes asociados a estas islas ha puesto de manifiesto que entre ellos se encuentran sobrerrepresentados genes implicados en el desarrollo (Illingworth *et al.*, 2008), así como genes específicos de la línea germinal (Weber *et al.*, 2007). Esto sugiere que, aunque existen otros mecanismos primarios de silenciamiento génico, la metilación del DNA constituye un mecanismo muy eficiente para mantener el silenciamiento de genes cuya expresión aberrante en un tejido inapropiado podría tener graves consecuencias.

En cualquier caso, tanto si la metilación ocurre como paso posterior a un silenciamiento mediado por otros mecanismos como si se produce de forma dirigida sobre determinados promotores de tejidos concretos, el modo por el que la célula distingue entre las islas que han de metilarse y las que no sigue siendo un misterio. El análisis de la secuencia de las islas CpG que se encuentran metiladas con respecto a las que se encuentran desprovistas de metilación de manera constitutiva no ha revelado ninguna diferencia significativa en términos de contenido en G+C y CpG O/E (Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Straussman *et al.*, 2009). Straussman *et al.* (2009) proponen que la ausencia de metilación se produce por el

reconocimiento motivos específicos en la secuencia del DNA. Sin embargo, esto no explica por qué hay islas que se metilan en unos tejidos pero no en otros.

2.2 Contribución de la transcripción activa a la protección de las islas CpG frente a la metilación

Como se recoge en el apartado 2.4 de la introducción (Pág. 22), la hipótesis más aceptada en la actualidad acerca de la protección de las islas CpG frente a la metilación apunta a que se produce por inhibición directa o indirecta de la unión de las DNMTs mediada por la maquinaria de inicio de la transcripción y por la estructura de la cromatina en las islas (Fig. 5C Pág. 24). El hecho de que la mayor parte de las islas CpG permanezcan libres de metilación incluso en aquellos tejidos en los que el gen asociado no se está expresando sugiere que el mecanismo que protege a estas regiones frente a la metilación actúa independientemente de la actividad transcripcional. No obstante, se ha observado que las islas CpG asociadas a genes de expresión específica de tejido tienden a metilarse en líneas celulares en cultivo, mientras que no ocurre lo mismo con islas asociadas a genes de expresión constitutiva (Antequera *et al.*, 1990). Esto indica que, pese a la mayor parte de las islas CpG se encuentran desprovistas de metilación, la ausencia de transcripción activa en las islas asociadas a genes específicos de tejido parece hacer a estas islas más susceptibles a la metilación, sugiriendo que la transcripción activa podría contribuir a que las islas se mantengan libres de metilación.

Por otra parte, experimentos de RNA-Seq han permitido observar la generación de transcritos bidireccionales de pequeño tamaño en el TSS de muchos de los promotores de humano y ratón, especialmente en aquellos promotores asociados a islas CpG (Core *et al.*, 2008; Seila *et al.*, 2008; Taft *et al.*, 2009). La generación de estos RNAs no muestra correlación

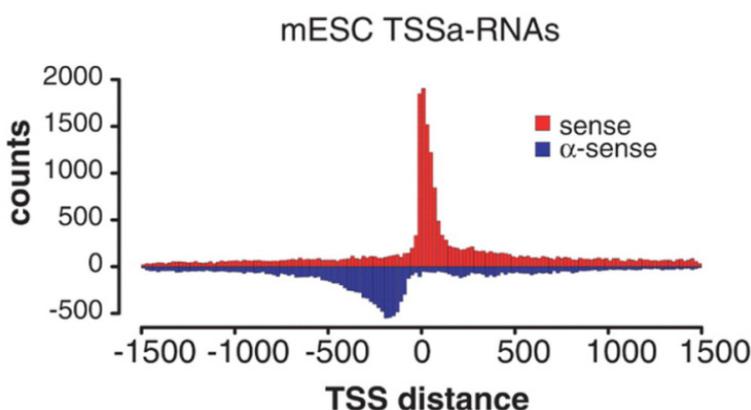


Figura 45. Distribución de RNAs pequeños asociados al TSS (TSSa-RNAs) en células ES de ratón. Extraído de Seila *et al.* (2008)

con el nivel de expresión de los genes asociados, lo que indica que la transcripción se inicia en muchos promotores, pero sólo en algunos se produce la elongación efectiva de los transcritos dando lugar a una RNA mensajero. Además, se ha comprobado que en ratón el 90% de los promotores en los que se detecta este fenómeno está enriquecido en

H3K4me3 y más del 50% presenta enriquecimiento en RNA Pol II (Seila *et al.*, 2008), ambas marcas características de las islas CpG no metiladas. Todo esto sugiere que, en contra de lo que se creía, el inicio de la transcripción en los promotores de las islas CpG podría contribuir activamente a que estas se mantengan libres de metilación, aunque esta transcripción no genere RNAs mensajeros.

Para comprobar si esta mayor susceptibilidad a la metilación en ausencia de transcripción activa era detectable no sólo en líneas celulares sino en muestras de tejido del organismo, analizamos el patrón de metilación de tres islas CpG asociadas a genes de expresión específica de tejido (*MYOD1*, *MT3* y *HBA1*) en DNA de sangre humana, donde ninguno de los tres genes seleccionados se expresa. Como control, analizamos el patrón de metilación de dos islas CpG asociadas a genes de expresión constitutiva (*TOP1* y *TIMM13*).

El análisis de estas islas mediante MeDIP sobre DNA de sangre y PCR con oligonucleótidos específicos sobre el inmunoprecipitado reveló que las islas asociadas a genes de expresión específica de tejido mostraban un enriquecimiento de en torno a diez veces con respecto al control negativo, mientras que las islas asociadas a genes de expresión constitutiva no mostraban enriquecimiento alguno (**Fig. 20 – 22** Pág. 55 y 56). Este enriquecimiento no se detectó cuando se realizó MeDIP sobre DNA de espermatozoides, en el que todas las islas analizadas mostraban una señal similar al control negativo (**Fig. 23** Pág. 57), lo que indicaba que los bajos niveles de metilación detectados en sangre eran significativos y no podían atribuirse a ruido de fondo de la técnica o a un sesgo por una amplificación más eficiente de estas regiones.

Para determinar si estos niveles de metilación podían vincularse a la ausencia de transcripción activa, realizamos MeDIP sobre DNA de la línea eritroide K562, en la que el gen *HBA1* se expresa de forma activa y pudimos comprobar cómo el enriquecimiento detectado en la isla asociada a este gen era equiparable al de la isla asociada a *TIMM13* (**Fig. 24** Pág. 58), confirmando que no se detectaba metilación cuando el gen se estaba expresando. Por otra parte, este experimento confirmó la tendencia de las islas asociadas a genes de expresión específica de tejido a metilarse en líneas celulares, ya que tanto la isla de *MYOD1* como la de *MT3* se encontraban totalmente metiladas en el DNA de K562.

Como se comenta en el apartado 2.4 de los resultados (Pág. 59), el enriquecimiento detectado mediante la técnica de MeDIP puede deberse a varios patrones de metilación. Este enriquecimiento podría deberse a que un bajo porcentaje de las células de la población tuvieran la región analizada densamente metilada, pero también se detectaría si todas las células de la población tuvieran unos pocos CpGs metilados en la región analizada. Con la

intención de determinar cuál podría ser el patrón de metilación de estas regiones analizamos en más detalle la isla CpG asociada al gen *HBA1*. Para ello, diseñamos una segunda pareja de oligonucleótidos que cubría la otra mitad de la isla CpG. Los resultados de este análisis mostraron un mayor enriquecimiento en el extremo 3' de la isla (**Fig. 25B** Pág. 60). Estos resultados parecían indicar que la isla podría estarse metilando por el extremo 3', haciendo así que se reduzca la ventana libre de metilación. Sería interesante determinar si en esta isla CpG, así como en otras islas asociadas a genes específicos de tejido, se produce la generación de transcritos de pequeño tamaño en tejidos en los que el gen no se expresa de forma activa. De ser así nos indicaría que el inicio de la transcripción en esos promotores podría contribuir a protegerlos frente a la metilación, pero la ausencia de elongación haría que la región protegida fuera de menor tamaño, lo que podría explicar que las islas empezaran a metilarse por su extremo 3', el más alejado del TSS. Este resultado concuerda con la observación de que las islas CpG son más proclives a metilarse por su extremo 3' en líneas celulares (Rideout *et al.*, 1994) (Cuadrado, Delgado & Antequera, sin publicar).

Intentamos caracterizar a nivel genómico este fenómeno por el que las islas CpG asociadas a genes específicos de tejido son más sensibles a la metilación en ausencia de transcripción activa combinando la inmunoprecipitación de DNA metilado con la hibridación de arrays de alta densidad (MeDIP-CHIP). Sin embargo, el análisis de los datos obtenidos nos permitió comprobar que la técnica de MeDIP-CHIP carece de la sensibilidad y resolución suficiente para estimar de forma fiable niveles bajos o intermedios de metilación (**Fig. 35 y 36** Pág. 69 y 70). No obstante, tal y como han puesto de manifiesto otros estudios en los que se combina el MeDIP con la hibridación de arrays de promotores (Rauch *et al.*, 2009; Straussman *et al.*, 2009), esta técnica resulta muy útil para la detección a nivel genómico de islas CpG metiladas (**Fig. 32** Pág 65).

Al observar el estado de metilación de las islas analizadas en nuestro estudio en los metilomas de líneas celulares de hESC, fibroblastos derivados de hESC y fibroblastos de neonato comprobamos que en los tres genes específicos de tejido se observa metilación en las islas CpG asociadas, mientras que las islas asociadas a los genes *TIMM13* y *TOP1* se encuentran totalmente desprovistas de metilación (**Fig. 46 y Fig. S3** Pág. 155-157). La metilación que se observa en las islas CpG asociadas a genes específicos de tejido se produce principalmente en el extremo 3' de las mismas, tal como sugieren nuestros datos.

La generación de metilomas sobre DNA de tejidos somáticos mediante el análisis de metilación por bisulfito sódico combinado con secuenciación masiva nos permitirá en un futuro determinar si esta mayor susceptibilidad a la metilación en ausencia de expresión activa del

gen asociado es extensible a todas las islas CpG asociadas a genes específicos de tejido. Además, contrastando estos datos de metilación a nivel genómico con la generación de transcritos de pequeño tamaño detectada en experimentos de RNA-seq (**Fig. 45**) podremos ver si existe una correlación entre el tamaño de la ventana libre de metilación en estos genes y la generación de estos transcritos incompletos y con ello, determinar la contribución del inicio de la transcripción a la protección de las islas CpG frente a la metilación.

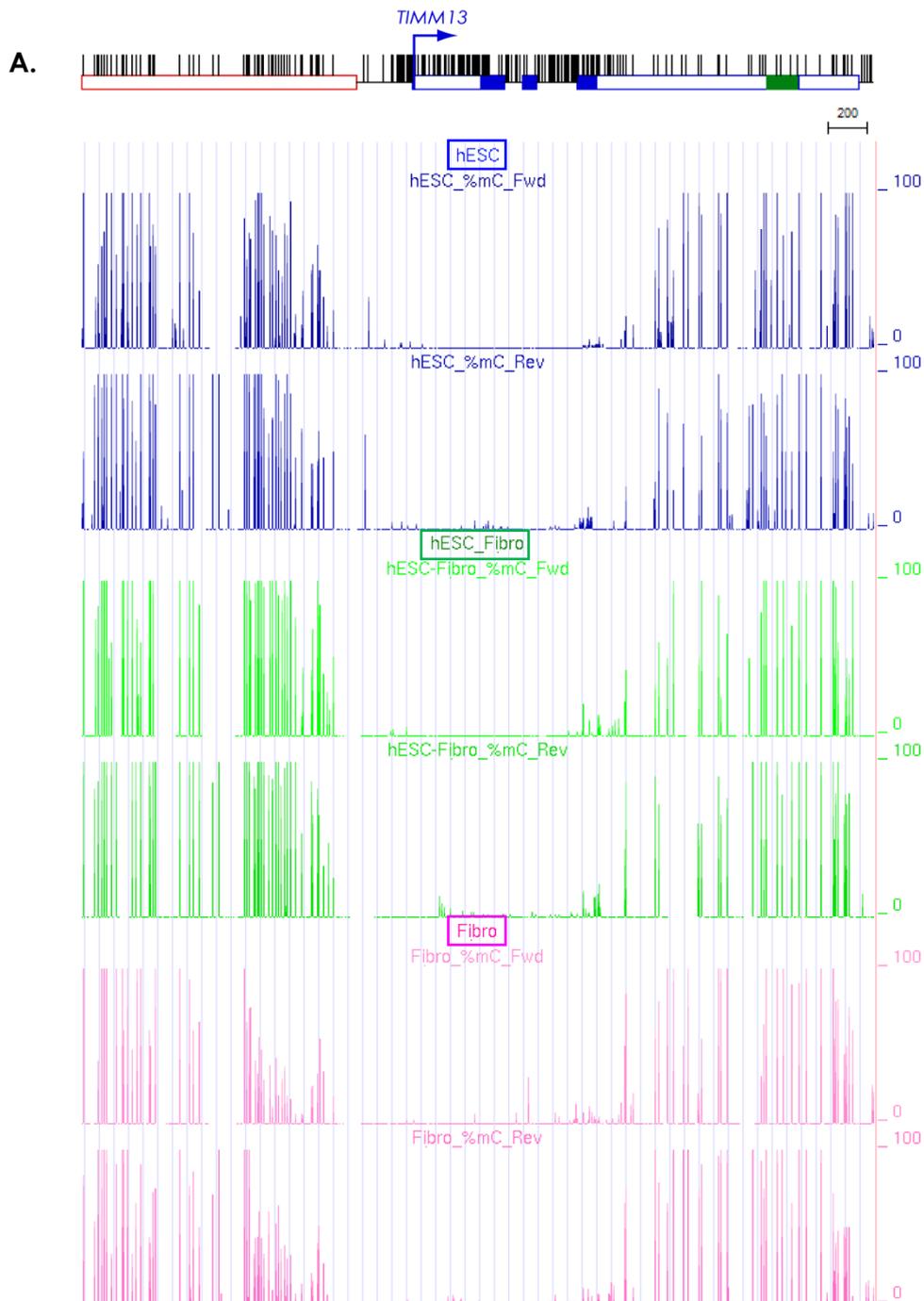


Figura 46. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para los genes *TIMM13* (A) y *HBA1* (B).

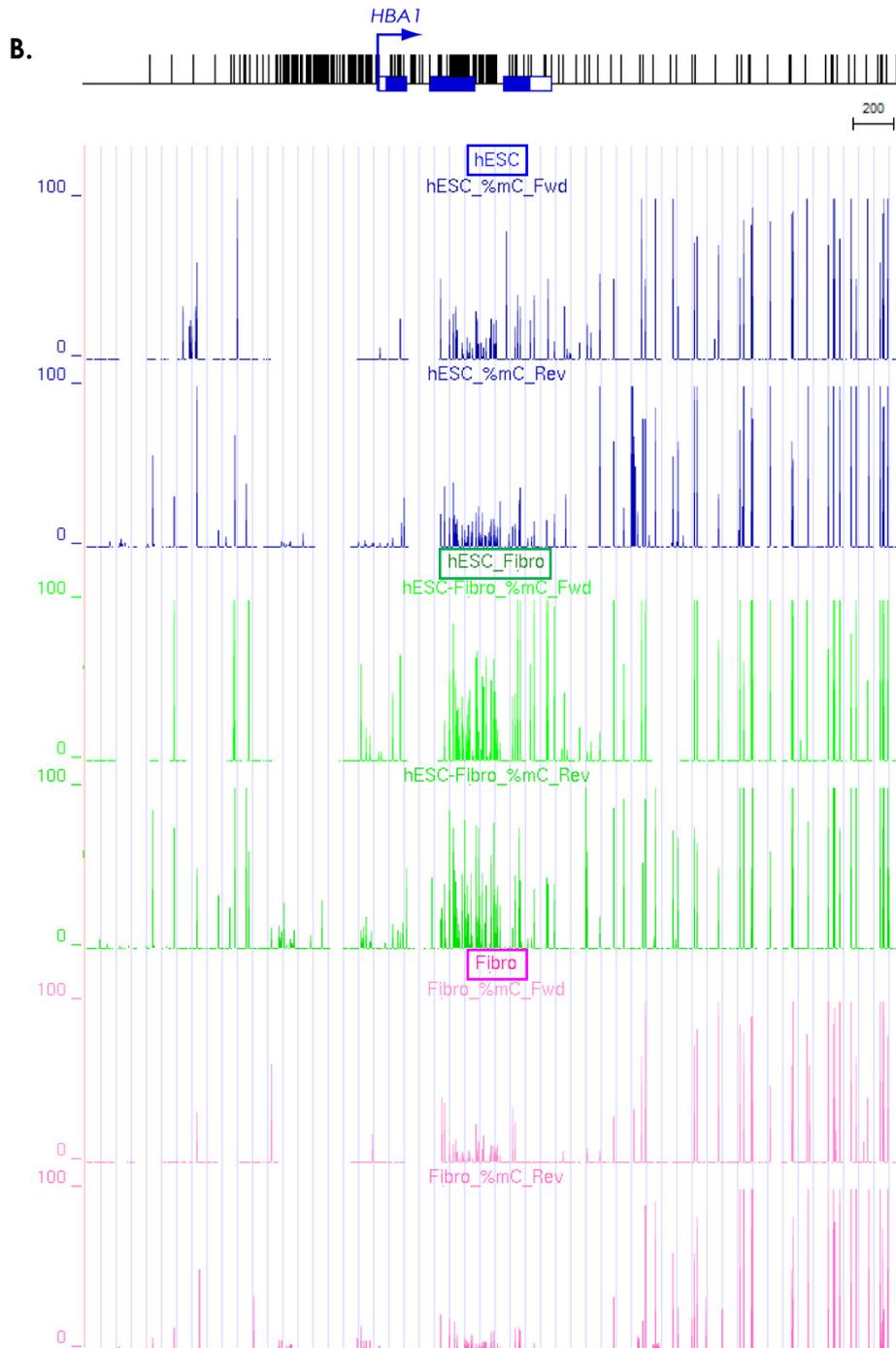


Figura 46. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para los genes *TIMM13* (A) y *HBA1* (B).

3. DINÁMICA DE GENERACIÓN DE LAS ISLAS CpG EN EL GENOMA DE MAMÍFEROS

La elevada colocalización de los promotores de los genes con las islas CpG apunta a que existe una relación funcional directa entre ambas características. No obstante, las islas CpG no constituyen elementos esenciales para la regulación de la expresión génica. Prueba de ello es que más del 30% de los genes de mamíferos se transcriben sin necesidad de tener una isla en su promotor. Además, la presencia de islas CpG en el promotor de los genes es una característica particular del genoma de vertebrados, especialmente en vertebrados de sangre caliente como aves y mamíferos. Por otra parte, las islas CpG son regiones altamente degeneradas en cuanto a secuencia, que no comparten entre ellas ninguna secuencia consenso, lo que es indicativo de que se trata de regiones dinámicas del genoma en las que la presión selectiva es menor que en las regiones codificantes de los genes. Estas características de las islas en cuanto a su localización y variabilidad de secuencia, nos llevaron a plantearnos la posibilidad de que las islas CpG hubieran surgido durante la evolución de los vertebrados como consecuencia de la inestabilidad asociada al inicio de la transcripción. Esta inestabilidad habría dejado una huella a modo de incremento en el contenido en G+C en los promotores de los genes.

Las islas CpG constituyen regiones con una composición de bases y unas características estructurales y epigenéticas que las diferencian del resto del genoma. Estas propiedades pudieron contribuir a facilitar el reclutamiento específico de los complejos transcripcionales a los promotores y, de este modo, haber supuesto una ventaja evolutiva importante a la hora de regular la expresión génica en genomas de gran tamaño como son los de los vertebrados.

3.1 Inestabilidad genética en los promotores asociados a islas CpG

Las islas CpG constituyen promotores desde los que se inicia la transcripción de los genes de mamíferos. Más del 70% de los genes humanos se encuentran asociados a una isla CpG, y este porcentaje podría ser muy superior, ya que estudios recientes han puesto de manifiesto que incluso las denominadas islas huérfanas presentan en su mayor parte propiedades características de promotores funcionales (Illingworth *et al.*, 2010; Maunakea *et al.*, 2010; Medvedeva *et al.*, 2010), lo que apunta a que podría tratarse de promotores de genes que no se han caracterizado, promotores alternativos de genes anotados o sitios de inicio de la generación de RNAs no codificantes. Los promotores asociados a islas CpG funcionan frecuentemente como promotores bidireccionales (Trinklein *et al.*, 2004) y es frecuente detectar

en ellos la generación de transcritos bidireccionales de pequeño tamaño, tal y como se ha indicado anteriormente (**Fig. 45** Pág. 92).

Por otra parte, las islas CpG funcionan como ORIs desde los que se inicia la replicación bidireccional del DNA. Aproximadamente un 50% de los ORIs descritos en humano colocalizan con islas CpG (Cadoret *et al.*, 2008; Sequeira-Mendes *et al.*, 2009) y se ha observado que los ORIs asociados a islas CpG son los que presentan una mayor eficiencia de activación (Sequeira-Mendes *et al.*, 2009). Un fenómeno similar al de la generación de RNAs de pequeño tamaño se ha descrito en el caso de los ORIs, en los que se detecta la acumulación de intermediarios de replicación de pequeño tamaño que coinciden con el TSS del gen adyacente durante la fase S del ciclo celular (Gómez & Antequera, 2008).

Las islas CpG son, por tanto, regiones del genoma sujetas a una gran actividad, lo que podría contribuir a que estas regiones estuvieran sometidas a una mayor inestabilidad, favoreciendo una mayor frecuencia en la aparición de daños en el DNA que deberían ser reparados. Una tendencia en la reparación de este daño hacia la introducción preferente de G o C iría generando en estas regiones un incremento en el contenido en G+C de las mismas que explicaría el sesgo en la composición de bases que presentan las islas con respecto al resto del genoma.

De acuerdo con esta hipótesis, se ha observado que existe una correlación positiva entre las tasas de sustitución y el nivel de expresión de los genes en la línea germinal (Mugal *et al.*, 2010). Además, los promotores con islas CpG presentan unas tasas de sustitución superiores a las detectadas para los promotores no asociados con islas, es decir, evolucionan más rápidamente (Carninci *et al.*, 2006). En relación con esto, se ha observado que genes que se expresan en la línea germinal o en el embrión poseen una isla CpG asociada a su promotor (Yoshikawa & Aizawa, 1988; Gardiner-Garden & Frommer, 1994; Daniels *et al.*, 1995; Daniels *et al.*, 1997; Ponger *et al.*, 2001). Este es el caso del gen de la α -globina, que se expresa en fases tempranas del desarrollo embrionario y posee una isla CpG en su promotor, mientras que otros genes, como el de la β -globina, carecen de isla CpG y no se expresan en el embrión ni en la línea germinal (Daniels *et al.*, 1997). Sería interesante establecer si existe también una correlación entre los ORIs funcionales en línea germinal y la presencia de islas CpG para determinar cómo afecta el inicio de la replicación en estas regiones a la inestabilidad de las mismas.

3.2 Incremento en el contenido en G+C en las regiones inestables del genoma

La hipótesis que hemos propuesto, por la que la inestabilidad genética en el entorno de las islas sería la responsable de su elevado contenido en guanina y citosina, predice que otras regiones del genoma sometidas a inestabilidad genética deberían presentar también un enriquecimiento en G+C y que debería existir una correlación entre el tamaño de la región inestable y la zona enriquecida en G+C.

Junto con el inicio de la transcripción y la replicación, otro de los procesos que genera inestabilidad en el genoma es la recombinación. La hipótesis de la conversión génica sesgada (BGC: *Biased Gene Conversion*) argumenta que las regiones sometidas a conversión génica sufren un sesgo en la reparación de los heterodúplex generados por la recombinación entre alelos heterocigóticos, que da lugar a la fijación preferente de alelos ricos en G+C (Galtier *et al.*, 2001). Esta hipótesis trata de explicar el elevado contenido en G+C de las isocoras, que son regiones del genoma cuyo tamaño llega a alcanzar las 300Kb, caracterizadas por un contenido en G+C superior al promedio genómico que se correlaciona positivamente con una mayor densidad de genes. De acuerdo con esta idea, numerosos estudios han detectado la existencia de una correlación positiva entre el contenido en G+C y las tasas de recombinación (Fullerton *et al.*, 2001; Meunier & Duret, 2004; Pollard *et al.*, 2006; Spencer *et al.*, 2006; Dreszer *et al.*, 2007; Duret & Arndt, 2008; Berglund *et al.*, 2009). Existe un amplio rango de tamaños de regiones sometidas a conversión génica que muestran un enriquecimiento en G+C, como el caso de los genes de la α -globina donde la región que sufre conversión génica ocupa 1 Kb (Fig. 47A) (Michelson & Orkin, 1983), o el caso de los genes de RNA ribosómico (Brock & Bird, 1997), donde el incremento en G+C se prolonga a lo largo de las 13,3 Kb de cada unidad transcripcional (Fig. 47B).



Figura 47. Contenido en G+C en regiones sometidas a conversión génica. A. Islas CpG asociada al promotor de los genes *HBA2* y *HBA1*. Se representa con una línea negra la región sometida a conversión génica.

B.

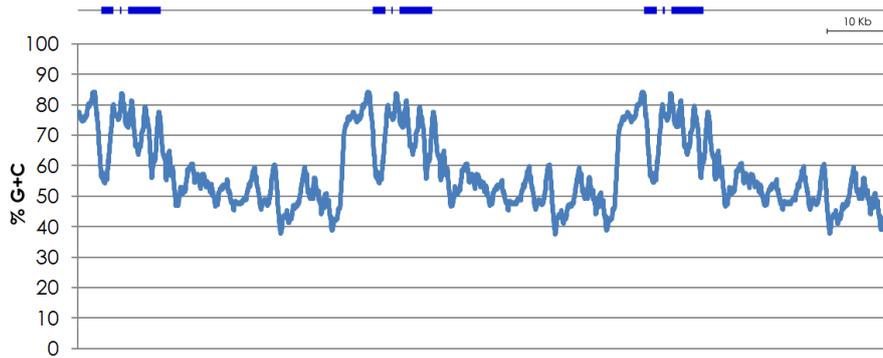


Figura 47. B. Genes del RNA ribosomal (18S, 5.8S y 28S). Se muestran tres repeticiones de la unidad transcripcional.

Otras regiones sometidas a conversión génica en las que se detecta un incremento en el contenido en G+C son las asociadas a los genes de choque térmico (Hsp70) (Kudla *et al.*, 2004), los genes del complejo de histocompatibilidad de clase II (Hogstrand & Bohme, 1999) y los *clusters* de genes de protocaderinas (Wu *et al.*, 2001).

La asociación entre las regiones ricas en G+C con procesos generadores de inestabilidad sugiere que el sesgo en la composición de bases de estas regiones es introducido por una o varias polimerasas implicadas en reparación. Entre todas las polimerasas implicadas en reparación en mamíferos hay una en particular que ha llamado nuestra atención. Se trata de pol λ , una polimerasa de la familia Pol X. Esta familia de proteínas participa en la reparación de roturas de doble cadena por el mecanismo de unión de extremos no homólogos (NHEJ), así como la eliminación de bases dañadas en el DNA mediante reparación por escisión de base (BER). Pol λ ha sido implicada, al menos *in vitro*, en ambos procesos de reparación (NHEJ: (Fiala *et al.*, 2004), (Nick McElhinny *et al.*, 2005), (Picher *et al.*, 2006), (Lee *et al.*, 2004); BER: (Fiala *et al.*, 2004), (García-Díaz *et al.*, 2002)). Además esta polimerasa carece de actividad exonucleasa 3'→5' (*proofreading*) (García-Díaz *et al.*, 2002) lo que la hace más proclive a cometer errores que otras DNA polimerasas. Ensayos *in vitro* han mostrado que la frecuencia con la que pol λ genera mutantes es seis veces mayor que la de pol β (Bebenek *et al.*, 2003), otra polimerasa que también pertenece a la familia X y que es la principal responsable de la reparación por BER. Estas evidencias, aunque se derivan de ensayos *in vitro*, muestran una tendencia de la pol λ a introducir errores a favor de C o G con mayor frecuencia que a favor de A o T (García-Díaz *et al.*, 2002; Bebenek *et al.*, 2003; Fiala *et al.*, 2004; Maga *et al.*, 2006; Picher *et al.*, 2006) lo cual es consistente con la posibilidad de que podría ser esta polimerasa la que, al reparar daños producidos en el DNA, introduciría el sesgo en la composición de bases que ha dado lugar a la formación de las islas CpG. Otro dato interesante acerca de esta familia de polimerasas es que está ausente en organismos como *Drosophila melanogaster* o *Caenorhabditis elegans*

(Uchiyama *et al.*, 2009), que carecen de islas CpG y en los que prácticamente no existe metilación del DNA (Simpson *et al.*, 1986; Tweedie *et al.*, 1997; Lyko *et al.*, 2000). Finalmente, se ha visto que la pol λ se sobreexpresa en testículo de ratón (García-Díaz *et al.*, 2000) así como en testículo y ovario humanos (Nagasawa *et al.*, 2000), detectándose un nivel basal en el resto de tejidos analizados. Esto tiene una gran importancia, puesto que para que el sesgo introducido en la secuencia de nucleótidos se acumule de forma hereditaria es requisito indispensable que tenga lugar en la línea germinal.

Existen diversas evidencias adicionales en la literatura de sesgos en la reparación del DNA hacia la introducción preferente de G o C en lugar de A o T (Brown & Jiricny, 1987; Bill *et al.*, 1998). Por ejemplo, al analizar el patrón de sustituciones en humanos con respecto a otros primates de aquellas regiones que han sufrido una evolución más acelerada con respecto al resto del genoma (HARs: *Human Accelerated Regions*) se observa que existe una tendencia a la sustitución de bases débiles, A y T, por bases fuertes, G y C (Pollard *et al.*, 2006; Berglund *et al.*, 2009; Katzman *et al.*, 2010) que también se ha observado en las islas CpG (Polak & Arndt, 2009; Polak *et al.*, 2010). Esta predisposición también se detecta cuando se analiza la fijación de sustituciones en poblaciones humanas mediante el estudio de polimorfismos (Webster & Smith, 2004; Katzman *et al.*, 2010).

Por otra parte, varios análisis han puesto de manifiesto que la composición de bases en torno al TSS en mamíferos es significativamente diferente de la composición de las regiones intergénicas y codificantes (Aerts *et al.*, 2004; Saxonov *et al.*, 2006). Se ha observado que tanto en humano como en ratón se produce un notable incremento en el contenido en G+C en torno al TSS. Este incremento es considerablemente mayor en los genes asociados a islas CpG, aunque también se observa una ligera diferencia en la composición del TSS en aquellos genes que no poseen una isla CpG en su promotor (**Fig. 48**). Esta colocación entre el TSS y el enriquecimiento en el contenido en G+C apunta nuevamente a la inestabilidad asociada al inicio de la transcripción como responsable de este sesgo en la composición de bases.

Otro ejemplo que ilustra la implicación de la inestabilidad genética en el incremento en G+C, es el caso del extremo 3' del gen *Fxy* tal y como se describió en el apartado 3.3 de resultados (Pág. 77). Esta región, tras ser incluida en el locus pseudoautosómico como consecuencia de una traslocación y debido, probablemente, a la elevada tasa de recombinación que tiene lugar en esta zona del cromosoma X, ha visto como se ha incrementado notablemente su contenido en G+C con respecto a la región ortóloga en otros organismos en los que dicha traslocación no ha tenido lugar (**Fig. 41** Pág. 79). Además, al

analizar el patrón de sustituciones se observa una clara tendencia a la sustitución de A y T por C y G (**Tabla 6** Pág.85).

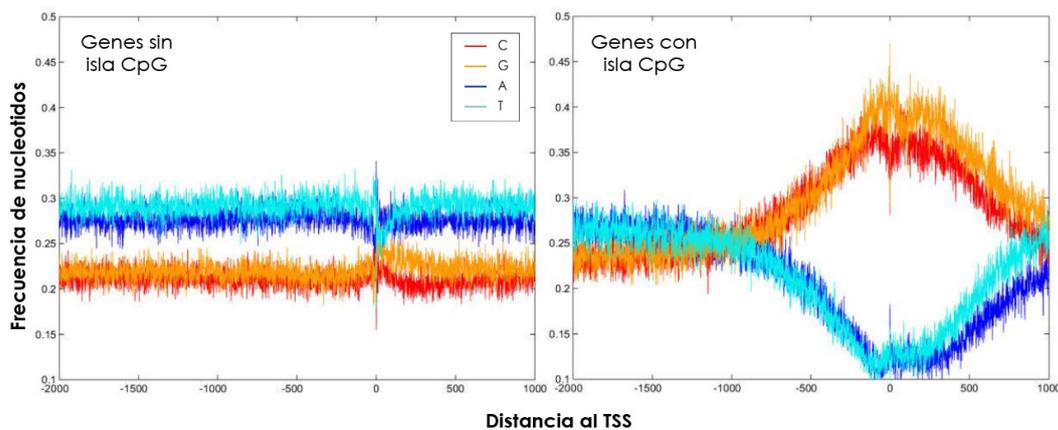


Figura 48. Análisis de la composición de bases en torno al TSS de los genes humanos. Se representa la frecuencia promedio para cada nucleótido en 750 genes con isla CpG y 750 genes sin isla CpG, alineados con respecto al TSS. (Adaptado de Aerts *et al.* 2004)

3.3 Dinámica de generación de las islas CpG en el genoma de mamíferos y mantenimiento del estado no metilado

Las evidencias recogidas en los apartados anteriores sustentan nuestra hipótesis de que el sesgo en la composición de bases que se detecta en las islas CpG, puede haberse generado como consecuencia de la inestabilidad que se produce en estas regiones, unida a un sesgo en la reparación del DNA hacia la introducción preferente de G y C. Esta hipótesis predice que, asumiendo que los mecanismos responsables de este incremento sigan actuando en la actualidad, deberíamos ser capaces de encontrar en el genoma islas CpG en distintos estados de formación.

De acuerdo con esta predicción, la región que más rápidamente ha evolucionado en humanos con respecto a otros primates (HAR1) se encuentra localizada en una isla CpG. La región ortóloga en ratón carece de isla CpG y tampoco muestra un contenido elevado en G+C, lo que apunta a que se trata de una isla que se ha generado con posterioridad a que los linajes de humano y mamífero divergieran en la evolución (**Fig. 49**).

Por otra parte, el análisis de las regiones no metiladas detectadas mediante experimentos de MeDIP-CHIP nos ha permitido encontrar regiones que, a pesar de no reunir los requisitos de secuencia necesarios para clasificarse como islas CpG poseen propiedades que se encuentran más próximas a las de las islas que al promedio genómico (**Tabla 5** Pág. 74), lo que sugiere que podría tratarse de islas que se están generando en el genoma humano.

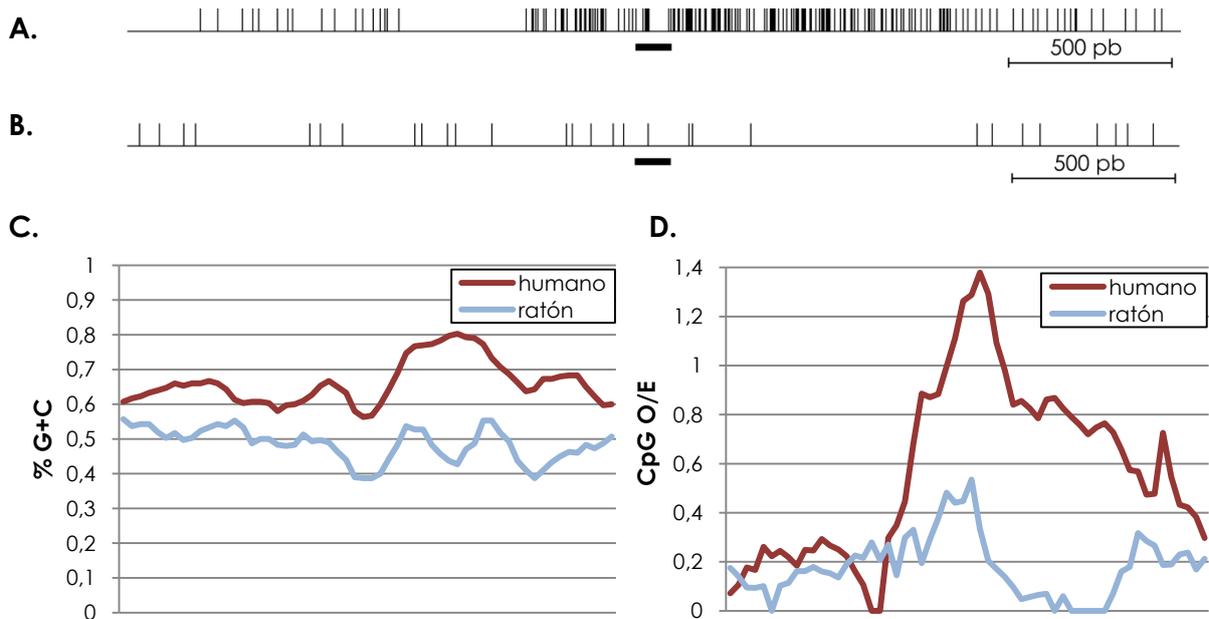


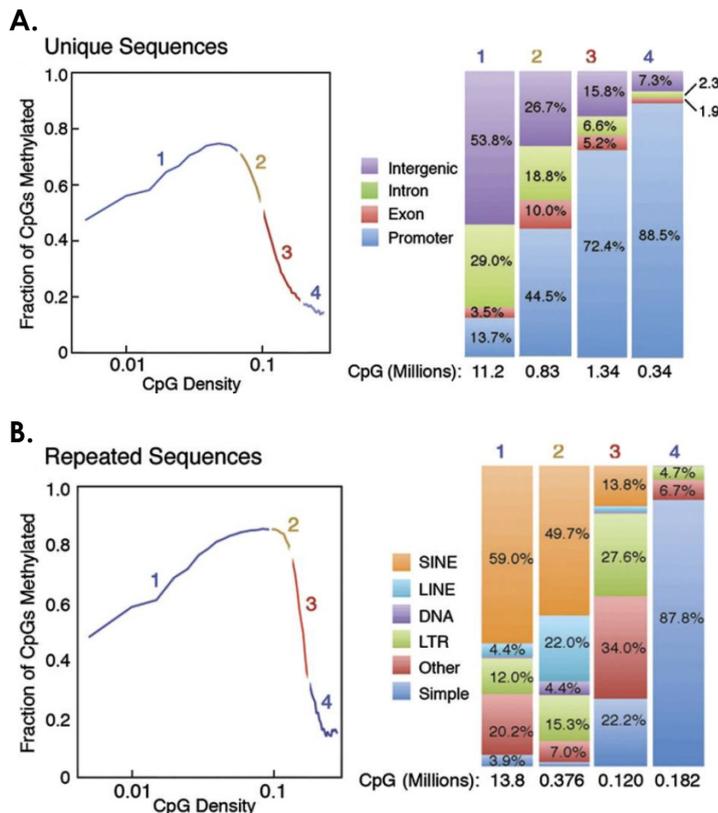
Figura 49. HAR1 (Human Accelerated Region 1). **A.** Mapa de la región en humano. Se representa con una línea negra la posición de la región HAR1 **B.** Mapa de la región en ratón. **C.** Contenido en G+C en humano y ratón. **D.** Relación de CpG O/E en humano y ratón.

El caso del gen *Fxy* en el locus pseudoautosómico en el genoma de *M. musculus* también podría incluirse como ejemplo de una región que está adquiriendo propiedades de isla CpG. En esta región no sólo se produce un enriquecimiento en G+C, sino que también se observa que, al menos en células ES esta región se encuentra parcialmente protegida de la metilación, lo cual concuerda con su elevado contenido en CpGs (Fig. 41B Pág. 79) y con la presencia de múltiples sitios de inicio de la transcripción a pesar de tratarse del extremo 3' del gen (Fig. 42 Pág. 81).

La inestabilidad genética y la reparación sesgada del daño en DNA pueden explicar la riqueza en G+C de las islas CpG, tal como indican las evidencias recogidas en el apartado anterior, pero no explica la elevada densidad de CpGs que presentan estas regiones con respecto al resto del genoma. Esta riqueza en CpGs se debe, probablemente, al hecho de que las islas CpG están libres de metilación y, por tanto, no están sometidas a la supresión de CpGs que se produce por la desaminación espontánea de las citosinas. Esto sugiere que el proceso responsable de la generación de las islas CpG debería contribuir al mantenimiento del estado no metilado de las mismas o, al menos, que la velocidad a la que se generan dinucleótidos CpG en estas regiones como consecuencia del incremento global de G+C tendría que ser superior a la velocidad a la que se pierden las citosinas metiladas por deaminación espontánea. Una posible explicación sería que estas regiones de inestabilidad comiencen a

incrementar su contenido en G+C hasta alcanzar un determinado umbral por encima del cual se produzca la exclusión de la metilación. Un elevado contenido en G+C podría facilitar la unión de factores de transcripción como SP1 y el reclutamiento de proteínas como CFP1 o KDM2A, que tienen afinidad por CpGs no metilados, y que contribuyeran al remodelamiento de la cromatina y la exclusión de las DNMTs de estas regiones.

De acuerdo con esta idea, al analizar el estado de metilación de 275 millones de CpGs en el genoma humano Edwards *et al.* (2010) observaron que el nivel de metilación aumentaba



de forma lineal a medida que se incrementaba la densidad de CpGs de las secuencias analizadas hasta alcanzar una densidad de CpGs de 0,025 (1CpG cada 40 nucleótidos) a partir de la cual descendía bruscamente, de modo que los niveles más bajos se detectaban en los promotores asociados a islas CpG (Edwards *et al.*, 2010) (Fig. 50A).

Sin embargo, este mismo estudio pone de manifiesto que la densidad de CpGs por sí sola no determina la ausencia de metilación de una región, puesto que en el caso de regiones repetitivas se observó que éstas se mantienen metiladas hasta densidades de CpG muy elevadas (Fig. 50B).

Figura 50. Relación entre el nivel de metilación y la densidad de CpGs. A. Regiones de copia única. **B.** Regiones repetitivas. (Extraído de Edwards *et al.* 2010).

Esta misma conclusión se deduce de los datos de metilación obtenidos para el exón 9 del gen Fxy de ratón (Fig. 43 Pág. 82), en los que se observa que existe cierta protección frente a la metilación en células ES, pero la misma región se encuentra densamente metilada en DNA procedente de cola. Por lo tanto, la riqueza en G+C y la elevada densidad de CpGs parecen constituir un requisito necesario pero no suficiente para determinar la ausencia de metilación de una región, como sugiere la existencia de islas que se metilen de forma diferencial en determinados tejidos (Yamada *et al.*, 2004; Weber *et al.*, 2005; Eckhardt *et al.*, 2006; Weber *et al.*, 2007; Illingworth *et al.*, 2008; Rakyan *et al.*, 2008; Rauch *et al.*, 2009; Straussman *et al.*, 2009; Illingworth *et al.*, 2010; Maunakea *et al.*, 2010).

No obstante, existen evidencias que apuntan a que el inicio de la transcripción podría ser el responsable no sólo del incremento en el contenido en G+C en torno al TSS (**Fig. 48** Pág. 102) sino de la protección de estas regiones frente a la metilación. En el apartado 2 de esta discusión (Pág. 91) se revisa la posible implicación de la transcripción en el mantenimiento del estado no metilado de las islas CpG. Asimismo, varios estudios han observado que la máxima relación de CpGs O/E coincide con el TSS y decrece a ambos lados del mismo (**Fig. 51**) (Saxonov *et al.*, 2006; Jiang *et al.*, 2007). Además, cuando se analiza la relación de dinucleótidos TpG y CpA, que se generan cuando se desaminan las citosinas metiladas, se observa que muestran la tendencia opuesta (Jiang *et al.*, 2007), lo que apunta a que existe una presión de metilación a ambos lados de las islas, mientras que es el centro de las mismas, asociado al TSS el que presenta una mayor protección frente a la metilación.

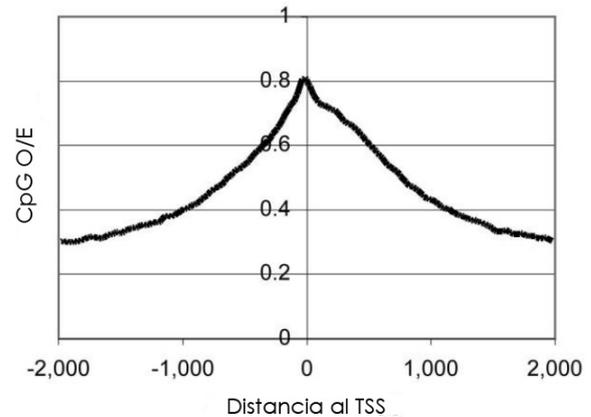


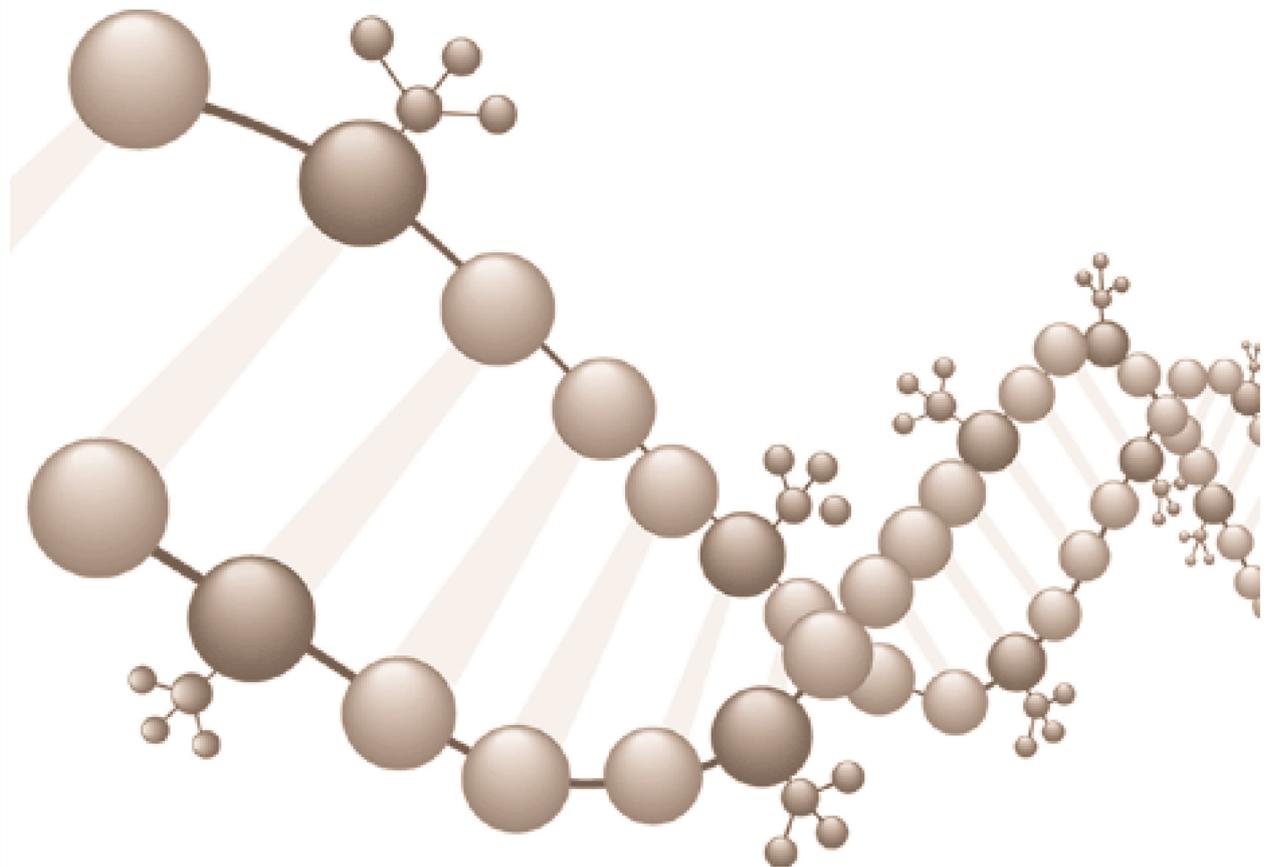
Figura 51. Relación de CpGs O/E promedio para 15880 genes humanos alineados con respecto al TSS. Se han considerado los genes incluidos en la base Ref Seq con TSS anotados. (Adaptado de Saxonov *et al.* 2005)

En conjunto, todos estos datos nos han llevado a proponer un modelo por el que el inicio de la transcripción en la línea germinal sería el principal mecanismo responsable del origen de las islas CpG en el genoma de mamíferos. Por una parte, se trata de una fuente de inestabilidad genética que, unida a un sesgo en la reparación del DNA habría dejado una huella a modo de incremento en el contenido en G+C en torno al TSS, y por otro lado, contribuiría a mantener estas regiones desprovistas de metilación evitando el silenciamiento de los genes asociados y la supresión de CpGs a la que está sometido el resto del genoma. Esta protección inicial frente a la metilación podría ser suficiente para que se generase una densidad de CpGs tal que permitiera la unión de CFP1 y KDM2A, provocando la remodelación de la cromatina en estas regiones y contribuyendo a una exclusión más efectiva de la metilación, lo que explicaría el brusco descenso en los niveles de metilación que se observa en regiones con una elevada densidad de CpGs (**Fig. 50A**).

El sesgo en la composición de bases en torno al TSS, que en un principio se produciría de forma accidental e inevitable como consecuencia de la inestabilidad asociada a esas regiones, pudo suponer una ventaja evolutiva importante para los vertebrados, facilitando a la maquinaria de transcripción la localización de los promotores en un genoma de gran tamaño. Esta maquinaria tendría que haberse adaptado a iniciar la transcripción en regiones con un

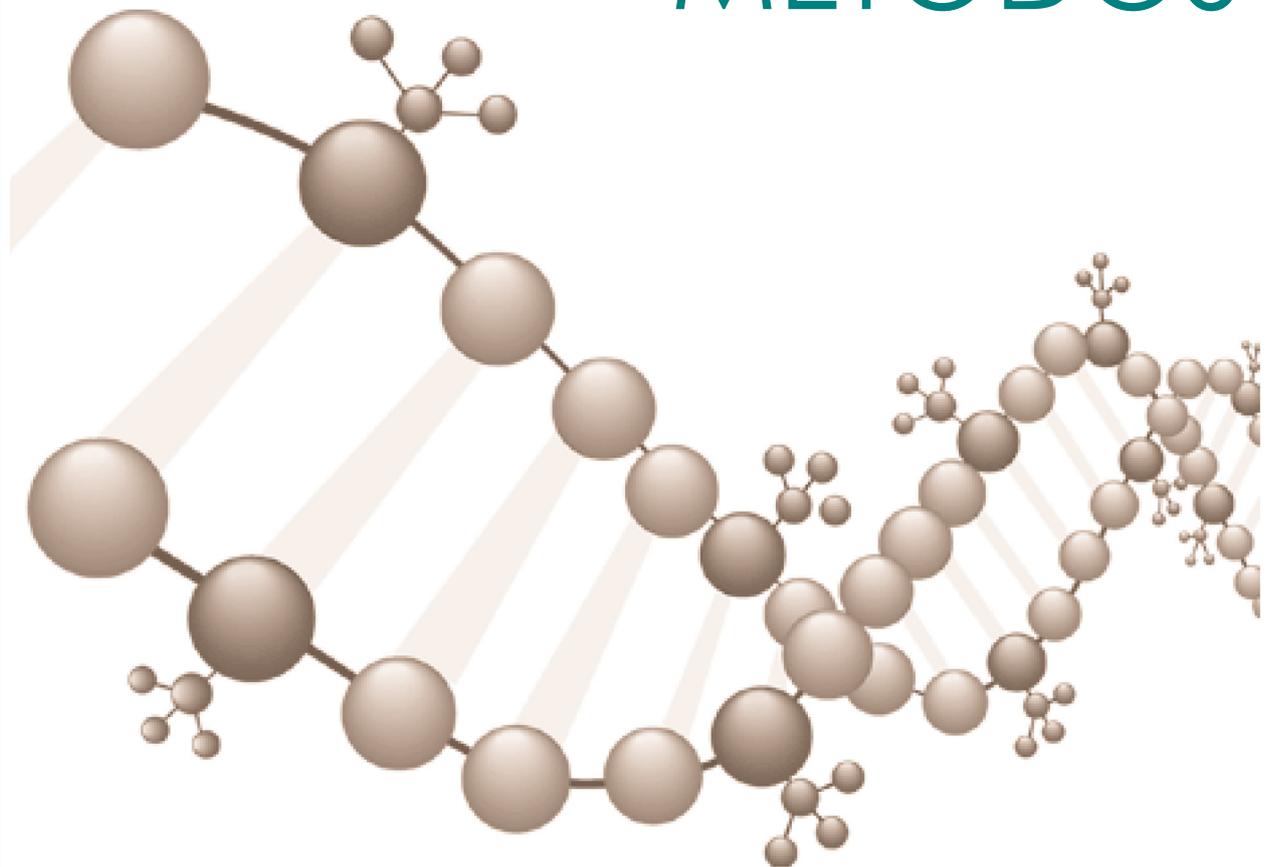
elevado contenido en G+C y eso explicaría por qué regiones cuyo incremento en la composición de guanina y citosina se ha generado como consecuencia de la inestabilidad asociada a la recombinación, como es el caso del extremo 3' del gen *Fxy* de ratón, ahora recluten la maquinaria de inicio de transcripción y sean capaces de excluir la metilación.

CONCLUSIONES



1. Es frecuente encontrar regiones conservadas adyacentes a islas CpG que, en la mayor parte de los casos analizados, establecen el límite de metilación en su extremo 5'. Sólo en algunos casos, el complejo de reconocimiento del origen (ORC) se une a estas regiones y determina que la replicación del DNA se inicie en ellas. Esto indica que el inicio de la replicación en las islas CpG no parece ser el mecanismo principal que establece el límite de la región no metilada, aunque no descarta que pueda contribuir parcialmente a su protección frente a la metilación.
2. Las islas CpG asociadas a genes específicos de tejido son más susceptibles de metilarse en tejidos somáticos normales en ausencia de transcripción activa, tal y como se había descrito previamente en líneas celulares. Esta metilación parece ocurrir direccionalmente, comenzando por su extremo 3', y reduce progresivamente la región libre de metilación.
3. Mediante inmunoprecipitación de DNA metilado e hibridación de *microarrays* hemos detectado la existencia de regiones desprovistas de metilación en el genoma humano con un contenido en G+C y una densidad de CpGs inferiores a los de las islas CpG. La caracterización de estas regiones en términos de secuencia e inicio de la transcripción sugiere que poseen propiedades más próximas a las de las islas CpG que a las del promedio genómico y que podría tratarse de islas en proceso de formación.
4. El extremo 3' del gen *Fxy* en la especie de ratón *Mus musculus* ha experimentado un incremento en su contenido en guanina y citosina con respecto a su región ortóloga en *Mus spretus*, rata y humanos. Este cambio está asociado a su traslocación a una región del cromosoma X sometida a una elevada tasa de recombinación, y sugiere que la inestabilidad genética resultante es la responsable del sesgo en su composición de bases. Además, esta región es capaz de excluir parcialmente la metilación en células ES, por lo que probablemente represente un estadio evolutivo intermedio en la formación de una isla CpG.
5. Generalizando a partir de las dos conclusiones anteriores, proponemos que las islas CpG asociadas a promotores podrían haberse generado como consecuencia de la inestabilidad genética asociada al inicio de la transcripción en la línea germinal. Por otra parte, regiones inestables del genoma no asociadas inicialmente a promotores habrían sufrido el mismo proceso de incremento en G+C. A partir de un determinado contenido en G+C, estas regiones serían capaces de reclutar la maquinaria de inicio de la transcripción y excluir la metilación, adquiriendo propiedades de islas CpG.

MATERIALES Y MÉTODOS



1. CULTIVOS CELULARES

Se utilizaron como modelo de células humanas las líneas HeLa (carcinoma de cérvix) y HEK 293 (embrionaria de riñón). Ambas líneas se incubaron en condiciones asépticas a 37 °C en una atmósfera humidificada al 5% de CO₂. Como medio de cultivo se utilizó DMEM suplementado con 10% de suero bovino fetal, 2 mM L-glutamina y 1% de penicilina-estreptomicina.

2. PURIFICACIÓN DE ÁCIDOS NUCLÉICOS

2.1 Extracción de DNA a partir de células en cultivo

La purificación de DNA genómico a partir de células en cultivo se llevó a cabo mediante procedimientos estándar (Sambrook *et al.*, 1989). Las células fueron lavadas dos veces con PBS a temperatura ambiente y a continuación se incubaron con tripsina durante 2-5 min a 37 °C. Se resuspendieron las células en tampón de lisis (50 mM Tris pH 8,0; 10 mM NaCl; 10 mM EDTA pH 8,0; 0,5% SDS; 100 µg/ml Proteinasa K) y se incubaron a 50°C durante 2h o a 37 °C durante toda la noche.

Se añadió al lisado un volumen de fenol y se mezcló suavemente. A continuación se centrifugó durante 10 min a 2500 g y se pasó la fase acuosa a un nuevo tubo. Se repitió el proceso utilizando un volumen de fenol:cloroformo:alcohol isoamílico (IAA) (25:24:1) y seguidamente un volumen de cloroformo:IAA (24:1). Se precipitó el DNA añadiendo dos volúmenes de etanol frío. Se realizó un lavado con etanol al 70% y se dejó secar a temperatura ambiente. Finalmente el DNA purificado fue resuspendido en TE (10 mM Tris pH 8,0; 1 mM EDTA).

2.2 Extracción de DNA de sangre

El primer paso para extraer DNA a partir de una muestra de sangre es separar las células mononucleadas del resto de componentes de la sangre. Para ello se centrifugaron 10 ml de sangre a 300 g durante 15 min a 4 °C. De este modo la sangre se separa en tres fases: una fase superior que corresponde al plasma, una interfase en la que se encuentran las células nucleadas y una fase inferior enriquecida en eritrocitos.

Se recuperó la interfase y se sometió a tres lavados con H₂O milliQ autoclavada. A continuación se resuspendieron las células en tampón de lisis I (50 mM Tris pH 8,0; 25 mM KCl; 5 mM MgCl₂; 8% sacarosa) y se centrifugó durante 10 min a 500 g para eliminar los restos de eritrocitos. Se resuspendieron las células en 2,5 ml de tampón de lisis II (2,125 ml de tampón de lisis I; 20 mM EDTA; Proteinasa K 80 µg/ml y 1% SDS) y se incubó a 55 °C durante toda la noche.

Se extrajo el lisado con fenol, fenol:cloroformo:IAA y cloroformo:IAA, como se describe en el apartado anterior, se precipitó con etanol frío y se resuspendió en TE.

2.3 Extracción de DNA de esperma

La extracción de DNA de esperma se llevó a cabo utilizando el método del tiocianato de guanidina (Hossain *et al.*, 1997).

Para separar los espermatozoides del líquido seminal se centrifugó el esperma durante 5 min a 1500 g y se lavó el precipitado 3 veces con PBS 1X. A continuación se resuspendió en un volumen pequeño de PBS y se observó al microscopio para determinar el número de espermatozoides así como el enriquecimiento de estos con respecto a otros tipos celulares presentes en el esperma. Se añadieron a continuación 5 ml de tampón de lisis por cada 3×10^6 - $1,2 \times 10^7$ espermatozoides y se incubó a 55 °C durante 3-4 horas.

Seguidamente se precipitó el DNA añadiendo dos volúmenes de isopropanol 100% frío y mezclando suavemente. Se formó un ovillo de DNA que se recuperó y se sometió a varios lavados con etanol al 70%, resuspendiéndolo finalmente en 100 µl de TE 1X.

Se cuantificó el DNA obtenido en Nanodrop, observando cuidadosamente la calidad del mismo (cocientes de absorbancia a 230 y 280) y, en las ocasiones en que fue necesario, se purificó el DNA con fenol cloroformo para eliminar posibles restos de proteínas.

2.4 Purificación de cadenas nacientes de DNA

La purificación de cadenas nacientes de DNA se llevó a cabo utilizando gradientes neutros de sacarosa (Abdurashidova *et al.*, 2000; Gómez & Antequera, 2008).

Para cada gradiente se empleo DNA genómico de aproximadamente 2×10^8 células creciendo exponencialmente. Este DNA se desnaturalizó durante 5-10 min a 100 °C y se fraccionó por tamaños mediante centrifugación en un gradiente discontinuo de sacarosa con siete capas de concentración creciente (de arriba abajo) del 5 al 20%, con una diferencia de 2,5% entre cada capa y la siguiente. Se centrifugó a 24000 rpm durante 20 h a 20 °C en un rotor Beckman SW-40Ti. Tras la centrifugación se recogieron 12 fracciones de 1ml cada una desde la superficie hasta el fondo. Se precipitó con etanol cada una de las fracciones y se tomaron alícuotas del DNA purificado para comprobar en un gel alcalino de agarosa al 1% (50 mM NaOH; 1 mM EDTA) que el fraccionamiento había sido correcto (**Fig.53**).

Las fracciones que contenían los intermediarios de replicación, cuyos rangos de tamaños son 100-600 pb, 300-800 pb y 600-2000 pb, fueron tratadas con polinucleotido-kinasa (PNK) para fosforilar los extremos 5'-OH. Esto permite que las moléculas de DNA que no tienen un primer de RNA en posición 5' sean digeridas por la λ -exonucleasa, de modo que eliminamos de estas fracciones todas aquellas moléculas pequeñas de DNA que no procedan directamente de intermediarios de replicación. Para la reacción de fosforilación se utilizaron 100 unidades de PNK en un volumen final de 200 μ l en presencia de 1 nM dATP. Se incubó durante 30 min a 37 °C y a continuación se inactivó la enzima añadiendo 6,25 μ g y 2,5 μ moles de EDTA en presencia de sarkosyl 0,125%. A continuación se purificó el DNA fosforilado mediante extracción y precipitación y se resuspendió en agua. El tratamiento con λ -exonucleasa se llevó a cabo en un volumen final de 80 μ l, utilizando 40 unidades de enzima e incubando a 37 °C durante toda la noche. Seguidamente se inactivó la enzima incubando la mezcla durante 10 min a 75 °C y se purificó nuevamente el DNA mediante extracción y precipitación, resuspendiéndolo finalmente en TE.

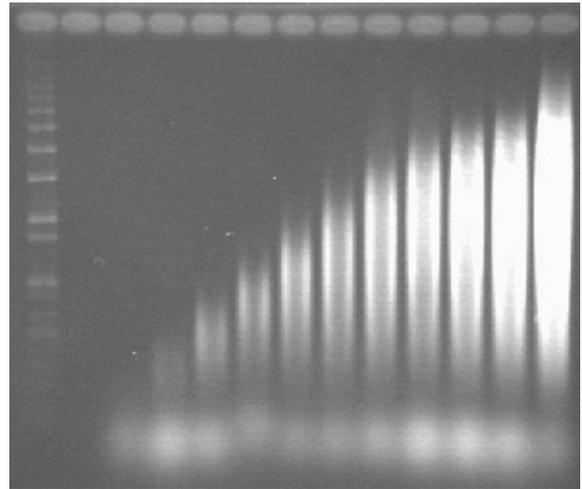


Figura 53. Fraccionamiento de DNA genómico en un gradiente neutro de sacarosa. Gel alcalino de agarosa al 1% en el que se muestran de derecha a izquierda las 12 fracciones recogidas del gradiente en orden de tamaño ascendente.

3. TÉCNICAS DE ANÁLISIS DE METILACIÓN DE DNA

3.1 Análisis de metilación con bisulfito sódico

La transformación del DNA mediante bisulfito sódico para el análisis de metilación es la más resolutive de las técnicas disponibles hasta el momento, ya que permite conocer el estado de metilación del DNA a nivel de nucleótido (Clark *et al.*, 1994; Rein *et al.*, 1997; Grunau *et al.*, 2001). El fundamento de esta técnica consiste en utilizar el bisulfito sódico para desaminar las citosinas transformándolas en uracilos en unas condiciones en las que las citosinas metiladas no son reactivas y, por lo tanto, no se ven modificadas. Esta diferente sensibilidad al bisulfito sódico entre citosinas metiladas y no metiladas es lo que nos va a permitir determinar el estado de metilación de cada dinucleótido CG (**Fig.54**).

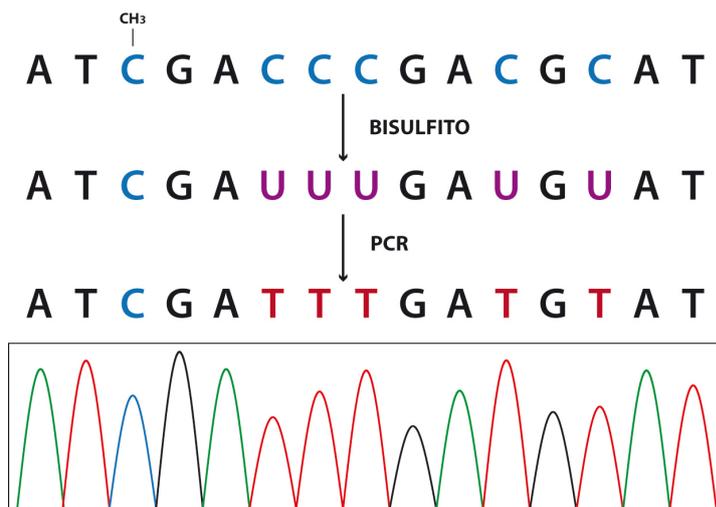


Figura 54. Análisis de metilación con bisulfito sódico. El tratamiento del DNA con bisulfito sódico transforma las citosinas no metiladas en uracilos. De este modo, si se analiza el DNA transformado mediante PCR con oligonucleótidos específicos para una región problema, se detectarán las citosinas no metiladas como timinas, mientras que sólo aquellas citosinas que se encontraban metiladas en el DNA de partida se mantendrán como tales en el DNA transformado.

3.1.1 Fragmentación del DNA genómico

Para facilitar la reacción de transformación del DNA con el bisulfito sódico se fragmentó el DNA genómico total mediante sonicación. También es posible utilizar una enzima de restricción para fragmentar el DNA. En este último caso será necesario comprobar que en la región que queremos analizar no exista ningún sitio de corte para la enzima utilizada.

3.1.2 Desnaturalización del DNA fragmentado

Se partió de una cantidad inicial de 5 a 10 µg de DNA en 20 µl de agua a los que se añadieron 30 µl de NaOH 1M, recién preparado, y 50 µl de agua. Esta mezcla se incubó durante 30 min a 37 °C para desnaturalizar el DNA fragmentado.

3.1.3 Transformación con bisulfito sódico

Para llevar a cabo la desaminación de las citosinas no metiladas y su consiguiente transformación en uracilos se añadió al DNA desnaturalizado 1ml de una solución de bisulfito sódico 3,6 M e hidroquinona 60 mM a pH 5,0. Se mezcló cuidadosamente por inversión y se incubó durante 16 horas a 50 °C añadiendo 300 µl de aceite mineral para evitar la evaporación.

A continuación se purificó el DNA transformado utilizando el kit Wizard®DNA Clean-Up System (Promega), siguiendo el protocolo especificado por el fabricante. Se recuperó el DNA en 100 µl de agua. Para evitar la formación de estructuras secundarias se añadieron 45 µl de NaOH 1 M y 5 µl de agua y se incubó a 37 °C durante 30 min.

Transcurrido este tiempo se añadieron 240 μl de acetato amónico 5 M pH 7,0 y 10 μl de agua, y se precipitó el DNA añadiendo 20 μg de glicógeno y dos volúmenes de etanol 100% frío. Se resuspendió el DNA precipitado en 100 μl de TE.

3.1.4 Análisis por PCR

Para analizar el estado de metilación de la región de interés se realizó una reacción de PCR con oligonucleótidos específicos para el DNA transformado por bisulfito. En un volumen final de 30 μl de reacción se añadieron: 1 a 5 μl del DNA tratado con bisulfito sódico como molde para la reacción de PCR, 1 μl de MgCl_2 25 mM, 2 μl de dNTPs 10 mM, 2 μl de cada oligonucleótido de un stock 50 μM , 0,25 μl de HotStarTaq® DNA polimerasa (Quiagen) de un stock a 5 U/ μl y 3 μl del tampón 10X específico de la polimerasa, completando con agua hasta 30 μl .

El programa de PCR utilizado fue el siguiente:

- 15 min a 95 °C para la activación de la HotStarTaq® DNA polimerasa
- 10 ciclos con las siguientes condiciones: 2 min de desnaturalización a 95 °C, 1 min a la temperatura de anillamiento específica para cada pareja de oligonucleótidos y 2 min de extensión a 68 °C.
- 20 ciclos en iguales condiciones de desnaturalización y anillamiento pero aumentando el paso de extensión a 68 °C 20 segundos en cada ciclo.
- Un paso de extensión final de 20 min a 68 °C.

Seguidamente se cortó la banda del gel correspondiente a nuestro producto de amplificación y se purificó el DNA utilizando el Illustra™MGFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare) según las especificaciones del fabricante. El tratamiento del DNA con bisulfito sódico produce una drástica reducción en la complejidad de la secuencia del mismo, haciendo que la PCR sea generalmente muy ineficiente. Debido a esto es frecuente que no se detecte el producto de la amplificación en un gel de agarosa. Cuando esto ocurre suele realizarse una segunda reacción de PCR utilizando como molde la primera y combinando uno de los oligonucleótidos de la reacción anterior con otro en posición más interna. El producto de esta segunda amplificación sí suele ser fácilmente detectable en gel de agarosa.

El producto de PCR purificado se cuantificó utilizando el NanoDrop® ND-1000 y se utilizaron 50-150 ng para la secuenciación automática del mismo o bien se clonó en el vector pGEM-T como paso previo a su secuenciación.

3.1.5 Clonación y obtención de DNA plasmídico

Se obtuvieron moléculas recombinantes mediante la ligación de los productos de PCR purificados con el vector plasmídico pGEM-T utilizando la T4 DNA ligasa (Fermentas). La mezcla de ligación se utilizó para transformar *E.coli* siguiendo el método descrito por Kushner *et al.* (1978) (Kushner, 1978) y a continuación se recuperaron las moléculas recombinantes por métodos estándar (Holmes & Quigley, 1981). La secuenciación automática de los fragmentos clonados se realizó en el servicio de secuenciación de la Universidad de Salamanca.

3.2 Inmunoprecipitación de DNA metilado (MeDIP)

La inmunoprecipitación de DNA metilado consiste en obtener, mediante la incubación con un anticuerpo que reconoce específicamente las 5-metil-citosinas (Eurogentec,(Reynaud *et al.*, 1992; Rougier *et al.*, 1998)), un inmunoprecipitado enriquecido en regiones de DNA metiladas (Weber *et al.*, 2005; Mohn *et al.*, 2009). La técnica consta de tres pasos básicos que son la sonicación, la desnaturalización y la incubación del DNA con el anticuerpo o inmunoprecipitación (**Fig. 55A**). A continuación pueden analizarse las regiones de interés mediante PCR en tiempo real o utilizar el inmunoprecipitado para hibridar un *microarray*.

3.2.1 Sonicación del DNA genómico

La sonicación se llevó a cabo en el sonicador Bioruptor® (Diagenode) en tubos de 1,5 ml. Se sonicaron 15 µg de DNA en 200 µl de TE 1X durante 10 min con pulsos de 30 segundos ON/30 segundos OFF a máxima potencia. A continuación se cargaron 500 ng en un gel de agarosa al 1% de concentración para comprobar que los fragmentos obtenidos tenían el tamaño apropiado, esto es, entre 300 y 1000 pares de bases (**Fig. 55B**).

3.2.2 Desnaturalización

Se diluyeron 4 µg del DNA sonicado en un volumen total de 450 µl de TE 1X (10 mM Tris-HCl pH 8,0 y 1 mM EDTA) y se procedió a su desnaturalización durante 10 min a 100 °C. Inmediatamente después se incubó en hielo durante 10 min.

3.2.3 Inmunoprecipitación de DNA metilado

Para cada inmunoprecipitación se añadieron al DNA desnaturalizado 51 µl de IP Buffer 10X (100 mM Na-Phosphate pH 7,0; 1,4 M NaCl; 0,5% Triton X-100) y 10 µl de anticuerpo y se incubó durante 2 horas a 4 °C con agitación. A continuación se añadieron a cada reacción 40

μ l de Dynabeads® (DynaL Biotech), que previamente se sometieron a dos lavados de 5 min con 800 μ l de PBS-BSA 0,1% a temperatura ambiente. Las Dynabeads® son lechos de poliestireno magnéticos que llevan unido covalentemente a su superficie un anticuerpo que reconoce específicamente las IgGs de ratón y que por tanto va a unirse al anticuerpo anti metilcitosinas, capturando así los complejos que este ha formado con los fragmentos de DNA metilado. Para recuperar las Dynabeads® tras cada lavado se utiliza un imán en el que se pueden adaptar los tubos de 1,5 ml. De este modo las Dynabeads® serán atraídas hacia la superficie del tubo en contacto con el imán y podremos retirar el sobrenadante sin perder Dynabeads®.

Tras la incubación con las Dynabeads® se realizaron tres lavados de 10 min de duración con IP Buffer 1X a temperatura ambiente. Para eluir el DNA inmunoprecipitado se añadieron a cada reacción 250 μ l de Digestion Buffer (50 mM Tris pH 8,0; 10 mM EDTA; 0,5% SDS) y 7 μ l de Proteinasa K a una concentración de 10 mg/ml y se incubó durante 3 horas a 50 °C con agitación. Seguidamente se realizó una extracción con fenol cloroformo y se precipitó el DNA con dos volúmenes de etanol 100%, 20 μ g de glicógeno y 20 μ l de NaCl 5 M, resuspendiendo el DNA precipitado en 60 μ l de TE 1X.

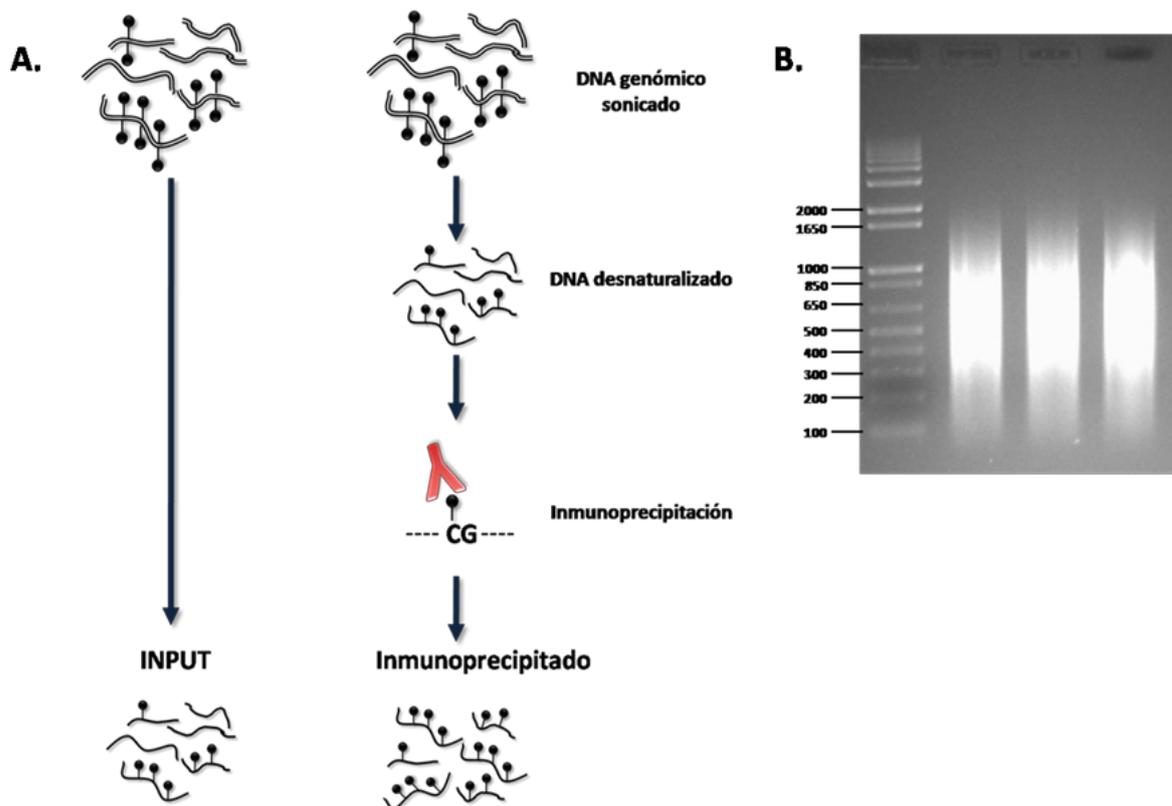


Figura 55. Inmunoprecipitación de DNA metilado (MeDIP). A. Esquema de los distintos pasos de los que consta la técnica de MeDIP. B. DNA genómico sonificado.

4. INMUNOPRECIPITACIÓN DE CROMATINA

La inmunoprecipitación de cromatina consiste en purificar aquellas regiones del genoma que interactúan directamente con una determinada proteína utilizando un anticuerpo que reconoce específicamente a dicha proteína (**Fig.56**). La inmunoprecipitación se llevó a cabo siguiendo el protocolo descrito por Frank *et al.* (2001) con algunas modificaciones (Frank *et al.*, 2001).

Las células se cultivaron en placas de 10 cm de diámetro en las condiciones anteriormente descritas. Es necesario que las células se estén dividiendo activamente de modo que podamos detectar más fácilmente las proteínas asociadas a los orígenes de replicación. Por ello partimos de placas con un 50-60% de confluencia.

4.1 Cross-linking

Se eliminó el medio de cultivo de cada placa y se añadieron 8 ml de PBS suplementado con PMSF. A continuación se agregó formaldehído a una concentración final del 1% y se incubó durante 10 min a temperatura ambiente con agitación suave. Transcurrido este tiempo se detuvo el *cross-linking* añadiendo glicina a una concentración final de 0,125 M e incubando durante 5 min más.

4.2 Extracción de los núcleos y eliminación del material citosólico

Se realizaron dos lavados con PBS frío y se recogieron las células en PBS suplementado con inhibidores de proteasas (PMSF 1 mM; aprotinina 1 µg/ml; leupeptina 1 µg /ml y pepstatina 1 µM) y se centrifugaron a 3200 g durante 5 min a 4 °C. Tras eliminar el sobrenadante se resuspendió el precipitado en el tampón de lisis I (5 mM PIPES pH 8,0; 85 mM KCl; 0,5% NP-40 e inhibidores de proteasas) y se incubó en hielo durante 10 min.

4.3 Lisis y fragmentación de la cromatina

Se centrifugó a 3200 g durante 5 min 4°C y se resuspendió el precipitado en el tampón de lisis II (50 mM Tris pH 8; 1% SDS; 10 mM EDTA pH 8,0 e inhibidores de proteasas) y se incubó en hielo durante 10 min.

Para fragmentar la cromatina se sonicó el lisado utilizando el Bioruptor (Diagenode) de manera que se obtuvieron fragmentos de un tamaño comprendido entre 100 y 800 pb. A continuación se centrifuga el lisado a 16200 g durante 10 min de modo que los restos nucleares

quedarán en el precipitado y la cromatina sonicada permanecerá en el sobrenadante. Se tomó una muestra de cromatina, se purificó el DNA y se cuantificó la cantidad de DNA utilizando el NanoDrop® ND-1000.

4.4 Inmunoprecipitación

Para cada inmunoprecipitación se utilizó el equivalente en cromatina a 100 µg de DNA, que fueron diluidos 10 veces con tampón de dilución IP (1% Triton X-100; 150 mM NaCl; 2 mM EDTA pH 8,0; 20 mM Tris pH 8,0 e inhibidores de proteasas).

Como paso previo a la incubación con el anticuerpo se lleva a cabo un prelavado de la cromatina añadiendo 80 µl de *beads* de agarosa unidas a proteína A/G (Santa Cruz Biotechnology) e incubando a 4 °C durante 2 h con agitación.

Se centrifugó brevemente (3 min a 400 g) para sedimentar las *beads*. Se recuperó el sobrenadante y se incubó con el anticuerpo deseado, utilizando 10 µg de los anticuerpos específicos para ORC2 y CDC6, y 2 µg para el anticuerpo específico para la histona H3. Se incubó a 4°C durante toda la noche.

Se recuperaron los complejos inmunes añadiendo 80 µl de *beads* de agarosa unidas a proteína A/G e incubando durante 2 h a 4 °C con agitación. A continuación se realizaron los siguientes lavados:

- Tres lavados con 1 ml de tampón *Mixed Micelle* (20 mM Tris pH 8,1; 150 mM NaCl; 5 mM EDTA pH 8,0; 5% sacarosa; 1% Triton X-100; 0,2% SDS y 0,2% NaN₃)
- Dos lavados con 1 ml de tampón 500 (50 mM HEPES pH 7,5; 0,1% ácido deoxicólico; 1% Triton X-100; 500 mM NaCl, 1 mM EDTA pH 8,0 y 0,2% NaN₃)
- Dos lavados con 1 ml de tampón de lavado LiCl/detergente (10 mM Tris pH 8,0; 0,5% ácido deoxicólico; 0,5% NP-40; 250 mM LiCl; 1 mM EDTA pH 8,0 y 0,2% NaN₃)
- Un lavado con 1 ml TE pH 8,0. Tras el último lavado se eluyeron las *beads* en 400 µl de tampón de elución (1% SDS y 100 mM NaHCO₃).

4.5 Reversión del *cross-linking* y purificación del DNA inmunoprecipitado

Para revertir el *cross-linking* se añadió NaCl a una concentración final de 200 mM y se incubó a 65 °C durante toda la noche. A continuación se incubó el inmunoprecipitado con tampón de digestión (10 mM EDTA pH 8,0; 40 mM Tris pH 6,5; Proteinasa K 50 µg/ml) a 45 °C durante 2-3 h con agitación. Finalmente cada muestra se extrajo con fenol/cloroformo, se

precipitó con 2 volúmenes de etanol y acetato sódico 0,3 M y se resuspendió en 50 µl de TE. Se utilizaron 2 µl para cada reacción de qPCR.

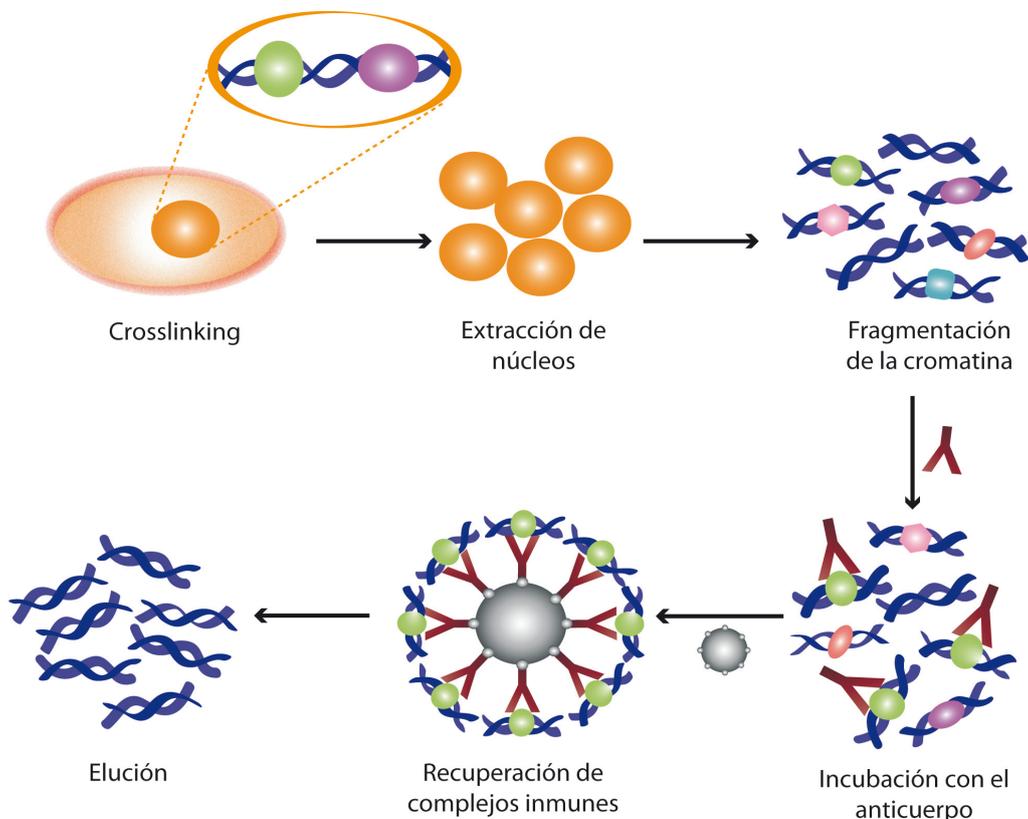


Figura 56. Inmunoprecipitación de cromatina. Esquema de los distintos pasos de los que consta la técnica de inmunoprecipitación de cromatina.

5. PCR CUANTITATIVA

Las reacciones de PCR cuantitativa en tiempo real (qPCR) se llevaron a cabo en el modelo ABI Prism 7000 Detection System (Applied Biosystems) usando el kit SYBR Premix EX Taq y siguiendo las instrucciones del fabricante. Todas las reacciones se realizaron por duplicado y los datos fueron analizados utilizando el ABI Prism 7000 SDS Software (versión 1.2.3, Applied Biosystems).

6. METILACIÓN DE DNA GENÓMICO “IN VITRO”

Para llevar a cabo la metilación *in vitro* del DNA genómico se utilizó la enzima SssI metiltransferasa (New England Biolabs), que metila la citosina de todos los dinucleótidos CpG, siguiendo el protocolo recomendado por el fabricante.

7. HIBRIDACIÓN Y ANÁLISIS DE TILING MICROARRAYS

7.1 Amplificación de DNA inmunoprecipitado

La amplificación de DNA inmunoprecipitado se llevó a cabo siguiendo las especificaciones del protocolo de Affymetrix para ensayos de inmunoprecipitación de cromatina (Fig.57).

El DNA inmunoprecipitado se desnaturalizó y se incubó con un oligonucleótido que posee un extremo 3' con 9 nucleótidos al azar que permitirán que se una a todos los fragmentos de DNA y un extremo 5' con una secuencia universal no degenerada que servirá para la posterior amplificación de los fragmentos con un oligonucleótido de secuencia complementaria. A continuación se realizaron 4 rondas de amplificación con Sequenasa para generar fragmentos de DNA de doble cadena y seguidamente se realizó una amplificación lineal utilizando un oligonucleótido complementario a la secuencia universal. En esta amplificación se incluye dUTP en la mezcla de dNTPs, lo cual permitirá la posterior fragmentación de las regiones amplificadas mediante las enzimas APE1 (*apurinic/apyrimidinic endonuclease*) y UDG (*uracil DNA glycosylase*) como paso previo a su marcaje a través de la enzima TdT (*terminal deoxynucleotidyl transferase*) e hibridación del array.

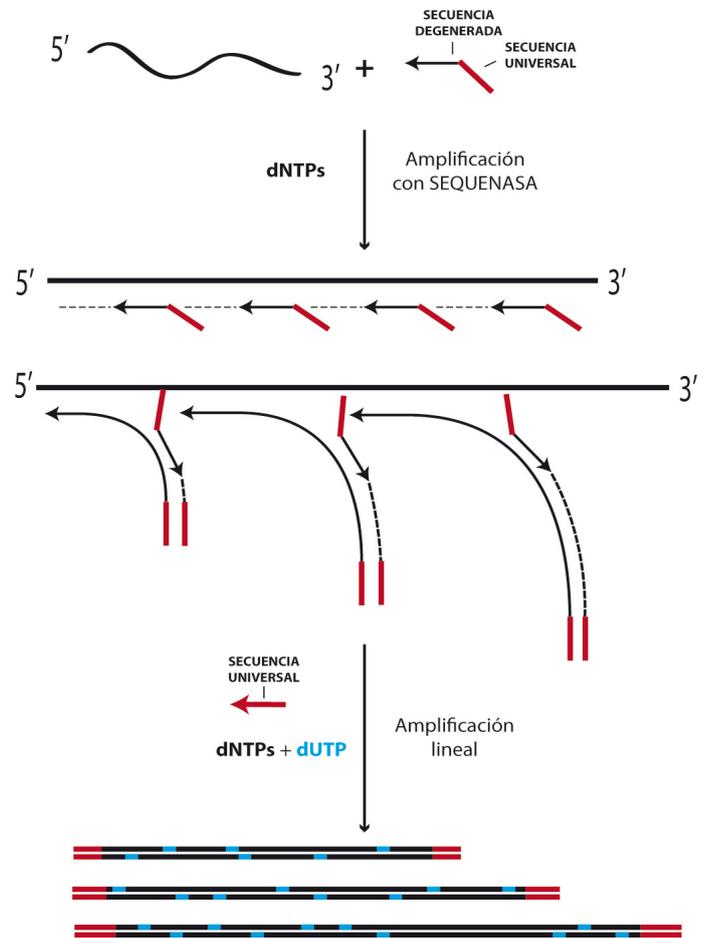


Figura 57. Amplificación de DNA inmunoprecipitado. Esquema de los distintos pasos de los que consta la amplificación del DNA inmunoprecipitado.

7.2 Fragmentación y marcaje del DNA amplificado

Tanto la fragmentación como el marcaje de los fragmentos amplificados se llevaron a cabo en la Unidad de genómica de la Universidad de Salamanca, utilizando el kit recomendado por el fabricante (*GeneChip WT Double-Stranded DNA Terminal Labeling Kit*).

7.3 Hibridación de los *tiling microarrays*

La hibridación de los *tiling microarrays* con el DNA inmunoprecipitado se llevó a cabo en la Unidad de genómica de la Universidad de Salamanca siguiendo las instrucciones del fabricante.

7.4 Análisis de datos

Los datos obtenidos de la hibridación de los *microarrays* fueron analizados utilizando el programa TAS (*Tiling Analysis Software*) de Affymetrix así como el IGB (*Integrated Genome Browser*) (Nicol *et al.*, 2009) con la colaboración del Dr. Luis Quintales, profesor titular del Departamento de Informática y Automática de la Universidad de Salamanca.

8. HERRAMIENTAS BIOINFORMÁTICAS

Una parte importante del trabajo de esta tesis doctoral se basa en la genómica comparativa mediante el análisis de alineamientos múltiples. Para ello hemos utilizado diversas herramientas bioinformáticas que se detallan a continuación.

8.1 Navegadores especializados

Para el rastreo de regiones conservadas en el extremo 5' de las islas CpG utilizamos principalmente dos navegadores, *Vista Browser* (Frazer *et al.*, 2004) y *UCSC Browser* (Kent *et al.*, 2002).

El *Vista Browser* es un navegador que incorpora alineamientos del genoma completo de muchos organismos de modo que permite visualizar el nivel de conservación de una región determinada de un genoma de referencia, definido por el usuario, con respecto a la región ortóloga de otras especies. Este programa permite establecer umbrales de conservación y tamaño para definir las regiones conservadas y además presenta distintas anotaciones que resultan de gran utilidad a la hora de analizar los patrones de conservación, como es el caso de la posición de los exones o las regiones repetitivas.

El UCSC *Browser* es un navegador que contiene múltiples anotaciones para los genomas de muchos organismos modelo. Entre las anotaciones disponibles existe una sección de genómica comparativa en la que podemos visualizar el grado de conservación de cada región con respecto a su ortóloga en distintos organismos. Una ventaja de este navegador con respecto al anterior es que nos permite ver el grado de conservación global de una región en un grupo de organismos (primates, mamíferos y vertebrados), mientras que el *Vista Browser* nos muestra el grado de conservación individual de cada especie con respecto al genoma de la especie de referencia. El grado de conservación se calcula en este caso mediante el programa PhastCons (Siepel *et al.*, 2005), que utiliza un método basado en modelos ocultos de Markov, donde no existen umbrales de tamaño o identidad de secuencia, sino que se calcula la probabilidad de cada nucleótido de pertenecer a un elemento conservado basándose en el alineamiento múltiple.

8.2 Programas de alineamientos múltiples

Los alineamientos múltiples se llevaron a cabo utilizando el programa ClustalX 2.0 (Chenna *et al.*, 2003), que utiliza un método de alineamiento global. Esto significa que las secuencias se alinean a lo largo de toda su longitud. En concreto los programas de la serie Clustal emplean un método de alineamiento global progresivo, para lo cual realizan múltiples alineamientos, comparando individualmente cada secuencia con todas las demás en alineamientos simples. Esto permite al programa generar un árbol filogenético de distancias entre las distintas secuencias en el que se basará para realizar el alineamiento múltiple.

Para el análisis de regiones conservadas en el extremo de las islas CpG utilizamos el programa Eshadow (Ovcharenko *et al.*, 2004), el cual tiene en cuenta no solo el grado de identidad de las secuencias, sino además la distancia filogenética entre ellas, permitiéndonos visualizar aquellas regiones de nuestro alineamiento que evolucionan más despacio, es decir, los elementos conservados. Paralelo el programa realiza en primer lugar un alineamiento múltiple utilizando el programa ClustalW y a continuación estima el porcentaje de divergencia en función del número de emparejamientos erróneos de cada secuencia con respecto al organismo base, que es definido por el usuario y en nuestro caso será siempre la secuencia humana. Con estos datos el programa genera una gráfica en la que se representa en el eje X la posición de la secuencia analizada y en el eje Y el porcentaje de divergencia. Sobre esta gráfica el programa nos permite establecer un umbral de tamaño mínimo y divergencia máxima para definir los elementos conservados.

8.3 Editor de alineamientos

Para la visualización y edición de los alineamientos múltiples realizados con el ClustalX hemos utilizado el programa Jalview 2.5.1 (Waterhouse *et al.*, 2009) que nos permite, entre otras cosas, colorear los alineamientos según el grado de identidad, lo que hará más sencilla la detección visual de las regiones conservadas.

8.4 Bases de datos

8.4.1 TiGER (*Tissue-specific Gene Expression and Regulation*)

TiGER contiene datos de expresión de alrededor de 20.000 genes en 30 tejidos humanos (Liu *et al.*, 2008). Los perfiles de expresión se basan en la cuantificación de ESTs (*Expressed Sequence Tags*), pequeñas fracciones de genes expresados obtenidas a partir de la secuenciación de cDNA clonado. Actualmente existen más de 8 millones de ESTs disponibles en la base de datos del NCBI para el genoma humano. El enriquecimiento (E) asignado a un gen X en un tejido Y se calcula como la relación entre los EST observados (EST_{Obs}) para el gen X en el tejido Y con respecto a los EST esperados (EST_{Esp}) en dicho tejido. El valor de EST esperados se calcula como el producto de los ESTs totales del gen X (EST_X) por la fracción de los ESTs totales del tejido Y (EST_Y) entre el número total de ESTs en los 30 tejidos analizados (EST_{TOTAL}).

$$E = EST_{Obs} / EST_{Esp} \quad EST_{Esp} = EST_X * (EST_Y / EST_{TOTAL})$$

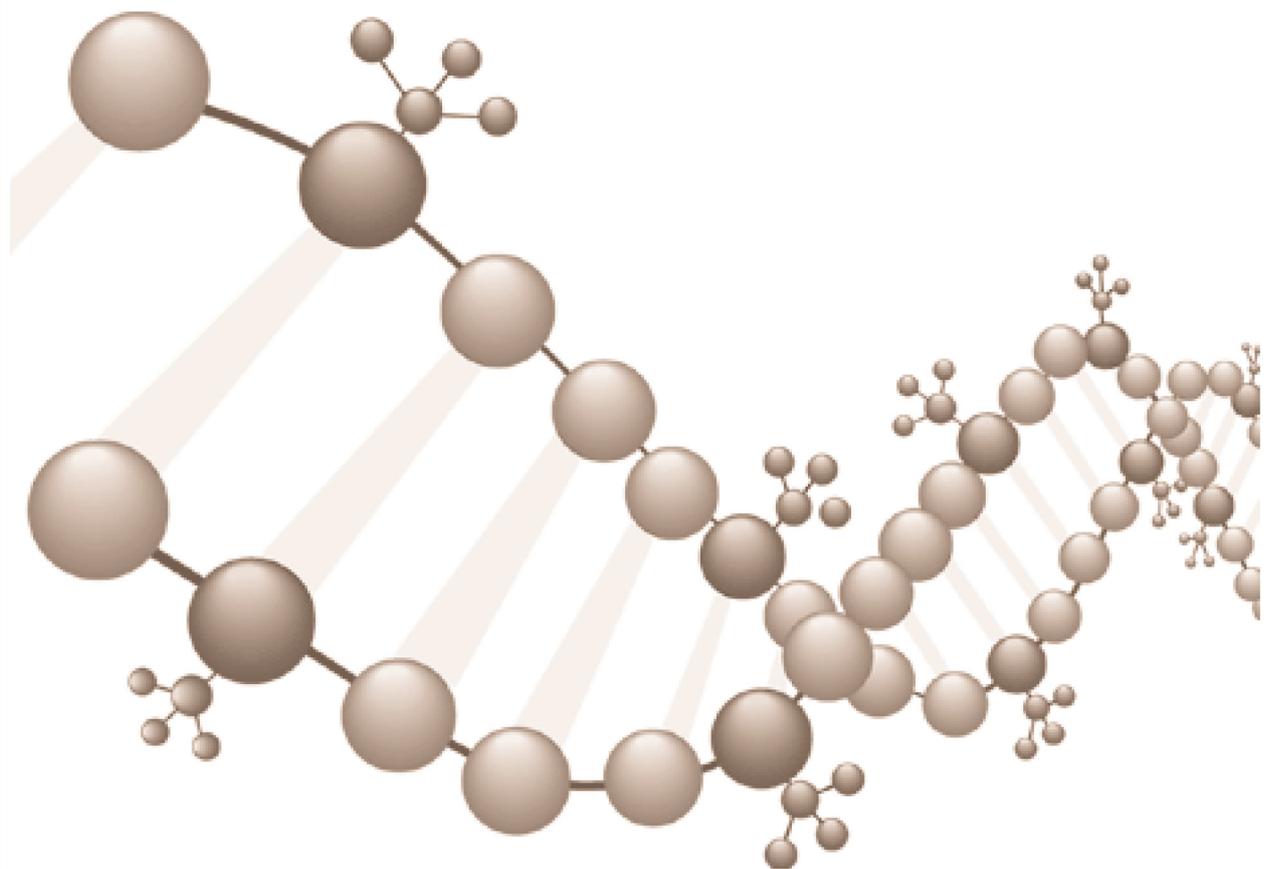
8.4.2 DBTSS (*Database of Transcriptional Start Sites*)

DBTSS contiene información precisa de la localización de los sitios de inicio de la transcripción (TSS) de los RNAs mensajeros (mRNAs) de humano y ratón (*M. musculus*) en varios tipos celulares (Suzuki *et al.*, 2002). El mapeo de los TSS se ha llevado a cabo mediante secuenciación masiva de librerías de cDNAs generadas a partir de RNAs mensajeros que mantienen intactos sus extremos 5' y 3'. Para ello se lleva a cabo un proceso de selección que consiste en sustituir la 7-metilguanosa (CAP) presente en el extremo 5' del mRNA por un oligonucleótido sintético. A continuación se combina este oligonucleótido con un oligodT, que se une específicamente a la cola poliA presente en el extremo 3' del mRNA. De este modo se garantiza que sólo se generará cDNA a partir de mRNAs intactos.

8.4.3 FANTOM (*Functional Annotation of the Mammalian Genome*)

FANTOM, al igual que DBTSS, contiene información acerca de la localización de los TSS de mRNAs de humano y ratón en distintos tipos celulares y tejidos (Kawaji *et al.*, 2009). La localización de los TSS se ha llevado a cabo mediante la técnica de CAGE (*Cap Analysis of Gene Expression*) (Shiraki *et al.*, 2003) que consiste en secuenciar fragmentos de 20 o 21 nucleótidos derivados del extremo 5' de los mRNAs. Para ello se lleva a cabo la unión de biotina a la estructura del CAP en el extremo 5' del mRNA y la introducción de un oligonucleótido sintético con sitios específicos para enzimas de restricción, que permite digerir el cDNA 20 o 21 nucleótidos corriente abajo del TSS generando así pequeños fragmentos derivados del extremo 5' del mRNA. A continuación estos fragmentos son clonados y secuenciados de forma masiva.

BIBLIOGRAFÍA



- Abdurashidova, G., Danailov, M.B., Ochem, A., Triolo, G., Djeliova, V., Radulescu, S., Vindigni, A., Riva, S., and Falaschi, A. (2003). Localization of proteins bound to a replication origin of human DNA along the cell cycle. *EMBO J* 22, 4294-4303.
- Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., and Falaschi, A. (2000). Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* 287, 2023-2026.
- Aerts, S., Thijs, G., Dabrowski, M., Moreau, Y., and De Moor, B. (2004). Comprehensive analysis of the base composition around the transcription start site in Metazoa. *BMC Genomics* 5, 34.
- Antequera, F. (2003). Structure, function and evolution of CpG island promoters. *Cellular and Molecular Life Sciences* 60, 1647-1658.
- Antequera, F. (2004). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J* 23, 4365-4370.
- Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 90, 11995-11999.
- Antequera, F., and Bird, A. (1999). CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr Biol* 9, R661-667.
- Antequera, F., Boyes, J., and Bird, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 62, 503-514.
- Aranda-Orgillés, B., Aigner, J., Kunath, M., Lurz, R., Schneider, R., and Schweiger, S. (2008). Active Transport of the Ubiquitin Ligase MID1 along the Microtubules Is Regulated by Protein Phosphatase 2A. *PLoS ONE* 3, e3507.
- Audit, B., Zaghloul, L., Vaillant, C., Chevereau, G., d'Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2009). Open chromatin encoded in DNA sequence is the signature of 'master' replication origins in human cells. *Nucleic Acids Res* 37, 6064-6075.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.
- Beall, E.L., Manak, J.R., Zhou, S., Bell, M., Lipsick, J.S., and Botchan, M.R. (2002). Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* 420, 833-837.
- Bebenek, K., Garcia-Díaz, M., Blanco, L., and Kunkel, T.A. (2003). The frameshift infidelity of human DNA polymerase lambda. Implications for function. *J Biol Chem* 278, 34685-34690.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2008). GenBank. *Nucleic Acids Research* 36, D25-D30.
- Berglund, J., Pollard, K.S., and Webster, M.T. (2009). Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol* 7, e26.
- Bill, C.A., Duran, W.A., Miselis, N.R., and Nickoloff, J.A. (1998). Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells. Competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. *Genetics* 149, 1935-1943.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6-21.
- Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91-99.

- Bird, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends Genet* 11, 94-100.
- Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* 38, 179-190.
- Borlado, L.R., and Méndez, J. (2008). CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis* 29, 237-243.
- Bosco, G., Du, W., and Orr-Weaver, T.L. (2001). DNA replication control through interaction of E2F-RB and the origin recognition complex. *Nat Cell Biol* 3, 289-295.
- Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96-99.
- Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science* 294, 2536-2539.
- Brock, G.J., and Bird, A. (1997). Mosaic methylation of the repeat unit of the human ribosomal RNA genes. *Hum Mol Genet* 6, 451-456.
- Brown, T.C., and Jiricny, J. (1987). A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. *Cell* 50, 945-950.
- Cadore, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A* 105, 15837-15842.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38, 626-635.
- Casillas, M.A., Lopatina, N., Andrews, L.G., and Tollefsbol, T.O. (2003). Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. *Molecular and Cellular Biochemistry* 252, 33-43.
- Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10, 295-304.
- Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 22, 2990-2997.
- Cohen, S.M., Brylawski, B.P., Cordeiro-Stone, M., and Kaufman, D.G. (2003). Same origins of DNA replication function on the active and inactive human X chromosomes. *J Cell Biochem* 88, 923-931.
- Colot, V., and Rossignol, J.L. (1999). Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21, 402-411.
- Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. *Nucleic Acids Res* 11, 647-658.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845-1848.
- Cuadrado, M., Sacristan, M., and Antequera, F. (2001). Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep* 2, 586-592.

- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* 31, 3497-3500.
- Daniels, R., Kinis, T., Serhal, P., and Monk, M. (1995). Expression of the myotonin protein kinase gene in preimplantation human embryos. *Human Molecular Genetics* 4, 389-393.
- Daniels, R., Lowell, S., Bolton, V., and Monk, M. (1997). Transcription of tissue-specific genes in human preimplantation embryos. *Hum Reprod* 12, 2251-2256.
- Delgado, S., Gomez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J* 17, 2426-2435.
- Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J Biol Chem* 280, 17986-17991.
- Dreszer, T.R., Wall, G.D., Haussler, D., and Pollard, K.S. (2007). Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. *Genome Res* 17, 1420-1430.
- Duret, L., and Arndt, P.F. (2008). The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* 4, e1000071.
- Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38, 1378-1385.
- Edwards, C.A., and Ferguson-Smith, A.C. (2007). Mechanisms regulating imprinted genes in clusters. *Current Opinion in Cell Biology* 19, 281-289.
- Edwards, J.R., O'Donnell, A.H., Rollins, R.A., Peckham, H.E., Lee, C., Milekic, M.H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., et al. (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res* 20, 972-980.
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8, 286-298.
- Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. (2008). Global organization of replication time zones of the mouse genome. *Genome Research* 18, 1562-1570.
- Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8, 188-194.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107, 8689-8694.
- Fiala, K.A., Abdel-Gawad, W., and Suo, Z. (2004). Pre-steady-state kinetic studies of the fidelity and mechanism of polymerization catalyzed by truncated human DNA polymerase lambda. *Biochemistry* 43, 6751-6762.
- Filippova, G.N. (2007). Genetics and Epigenetics of the Multifunctional Protein CTCF. In *Current Topics in Developmental Biology*, P.S. Gerald, ed. (Academic Press), pp. 337-360.

- Fraga, M.F., Agrelo, R., and Esteller, M. (2007). Cross-Talk between Aging and Cancer. *Annals of the New York Academy of Sciences* 1100, 60-74.
- Fraga, M.F., and Esteller, M. (2007). Epigenetics and aging: the targets and the marks. *Trends in Genetics* 23, 413-418.
- Frank, S.R., Schroeder, M., Fernandez, P., Taubert, S., and Amati, B. (2001). Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes & Development* 15, 2069-2082.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32, W273-279.
- Fullerton, S.M., Bernardo Carvalho, A., and Clark, A.G. (2001). Local rates of recombination are positively correlated with GC content in the human genome. *Mol Biol Evol* 18, 1139-1142.
- Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. (2001). GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159, 907-911.
- Garcia-Diaz, M., Bebenek, K., Sabariego, R., Dominguez, O., Rodriguez, J., Kirchhoff, T., Garcia-Palomero, E., Picher, A.J., Juarez, R., Ruiz, J.F., et al. (2002). DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J Biol Chem* 277, 13184-13191.
- Garcia-Diaz, M., Dominguez, O., Lopez-Fernandez, L.A., de Lera, L.T., Saniger, M.L., Ruiz, J.F., Parraga, M., Garcia-Ortiz, M.J., Kirchhoff, T., del Mazo, J., et al. (2000). DNA polymerase lambda (Pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J Mol Biol* 301, 851-867.
- Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Biology* 196, 261-282.
- Gardiner-Garden, M., and Frommer, M. (1994). Transcripts and CpG islands associated with the pro-opiomelanocortin gene and other neurally expressed genes. *J Mol Endocrinol* 12, 365-382.
- Geiman, T.M., Sankpal, U.T., Robertson, A.K., Zhao, Y., and Robertson, K.D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem Biophys Res Commun* 318, 544-555.
- Ghazi, H., Gonzales, F.A., and Jones, P.A. (1992). Methylation of CpG-island-containing genes in human sperm, fetal and adult tissues. *Gene* 114, 203-210.
- Ghosh, M., Kemp, M., Liu, G., Ritzl, M., Schepers, A., and Leffak, M. (2006). Differential binding of replication proteins across the human c-myc replicator. *Mol Cell Biol* 26, 5270-5283.
- Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S., et al. (1994). Fine mapping of a replication origin of human DNA. *Proc Natl Acad Sci U S A* 91, 7119-7123.
- Glass, J.L., Thompson, R.F., Khulan, B., Figueroa, M.E., Olivier, E.N., Oakley, E.J., Van Zant, G., Bouhassira, E.E., Melnick, A., Golden, A., et al. (2007). CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids Research* 35, 6798-6807.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74, 481-514.
- Gómez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. *Genes & Development* 22, 375-385.

- Gomez, M., and Brockdorff, N. (2004). Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage. *Proc Natl Acad Sci U S A* 101, 6923-6928.
- Gonzalez, S., Klatt, P., Delgado, S., Conde, E., Lopez-Rios, F., Sanchez-Cespedes, M., Mendez, J., Antequera, F., and Serrano, M. (2006). Oncogenic activity of CDC6 through repression of the INK4/ARF locus. *Nature* 440, 702-706.
- Grunau, C., Clark, S.J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* 29, E65-65.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77-88.
- Hackenberg, M., Previti, C., Luque-Escamilla, P., Carpena, P., Martinez-Aroza, J., and Oliver, J. (2006). CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics* 7, 446.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 452, 877-881.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78-82.
- Han, L., and Zhao, Z. (2009). CpG islands or CpG clusters: how to identify functional GC-rich regions in a genome? *BMC Bioinformatics* 10, 65.
- Hogstrand, K., and Bohme, J. (1999). Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. *Immunogenetics* 49, 446-455.
- Holmes, D.S., and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* 114, 193-197.
- Holliday, R., and Grigg, G.W. (1993). DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 285, 61-67.
- Hossain, A.M., Rizk, B., Behzadian, A., and Thorneycroft, I.H. (1997). Modified guanidinium thiocyanate method for human sperm DNA isolation. *Mol Hum Reprod* 3, 953-956.
- Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PLoS Biol* 6, e22.
- Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. *FEBS Lett* 583, 1713-1720.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 6.
- Jiang, C., Han, L., Su, B., Li, W.H., and Zhao, Z. (2007). Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. *Mol Biol Evol* 24, 1991-2000.
- Jones, P.A., and Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10, 805-811.

- Karnani, N., Taylor, C.M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* 21, 393-404.
- Karpf, A.R., and Matsui, S. (2005). Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res* 65, 8635-8639.
- Katzman, S., Kern, A.D., Pollard, K.S., Salama, S.R., and Haussler, D. (2010). GC-biased evolution near human accelerated regions. *PLoS Genet* 6, e1000960.
- Kawaji, H., Severin, J., Lizio, M., Waterhouse, A., Katayama, S., Irvine, K., Hume, D., Forrest, A., Suzuki, H., Carninci, P., et al. (2009). The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Genome Biology* 10, R40.
- Keller, C., Ladenburger, E.M., Kremer, M., and Knippers, R. (2002). The origin recognition complex marks a replication origin in the human TOP1 gene promoter. *J Biol Chem* 277, 31430-31440.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome Browser at UCSC. *Genome Res* 12, 996-1006.
- Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.
- Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol Cell Biol* 25, 4388-4396.
- Kreitz, S., Ritzi, M., Baack, M., and Knippers, R. (2001). The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells. *J Biol Chem* 276, 6337-6342.
- Kudla, G., Helwak, A., and Lipinski, L. (2004). Gene conversion and GC-content evolution in mammalian Hsp70. *Mol Biol Evol* 21, 1438-1444.
- Kushner, S. (1978). An improve method for transformation Escherichia coli with ColE1 derived plasmids. In *Genetic engineering*, H.W.B.a.S. Nicosia, ed. (Amsterdam, Elsevier/North-Holland Biomedical Press), pp. 17-23.
- Ladenburger, E.-M., Keller, C., and Knippers, R. (2002). Identification of a Binding Region for Human Origin Recognition Complex Proteins 1 and 2 That Coincides with an Origin of DNA Replication. *Mol Cell Biol* 22, 1036-1048.
- Laird, P.W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 11, 191-203.
- Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics* 13, 1095-1107.
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res* 20, 320-331.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11, 204-220.
- Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem* 276, 44669-44676.

- Lee, J.W., Blanco, L., Zhou, T., Garcia-Diaz, M., Bebenek, K., Kunkel, T.A., Wang, Z., and Povirk, L.F. (2004). Implication of DNA polymerase lambda in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts. *J Biol Chem* 279, 805-811.
- Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322.
- Liu, X., Yu, X., Zack, D., Zhu, H., and Qian, J. (2008). TiGER: A database for tissue-specific gene expression and regulation. *BMC Bioinformatics* 9, 271.
- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538-540.
- Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev* 8, 2282-2292.
- Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A., and Issa, J.-P.J. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Research* 20, 332-340.
- Maga, G., Shevelev, I., Villani, G., Spadari, S., and Hubscher, U. (2006). Human replication protein A can suppress the intrinsic in vitro mutator phenotype of human DNA polymerase lambda. *Nucleic Acids Res* 34, 1405-1415.
- Margulies, E.H., Cooper, G.M., Asimenos, G., Thomas, D.J., Dewey, C.N., Siepel, A., Birney, E., Keefe, D., Schwartz, A.S., Hou, M., et al. (2007). Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res* 17, 760-774.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253-257.
- Medvedeva, Y.A., Fridman, M.V., Oparina, N.J., Malko, D.B., Ermakova, E.O., Kulakovskiy, I.V., Heinzl, A., and Makeev, V.J. (2010). Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics* 11, 48.
- Mesner, L.D., and Hamlin, J.L. (2005). Specific signals at the 3' end of the DHFR gene define one boundary of the downstream origin of replication. *Genes & Development* 19, 1053-1066.
- Meunier, J., and Duret, L. (2004). Recombination drives the evolution of GC-content in the human genome. *Mol Biol Evol* 21, 984-990.
- Michelson, A.M., and Orkin, S.H. (1983). Boundaries of gene conversion within the duplicated human alpha-globin genes. Concerted evolution by segmental recombination. *J Biol Chem* 258, 15245-15254.
- Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG Mutability and Tumorigenesis in MBD4-Deficient Mice. *Science* 297, 403-405.
- Minami, H., Takahashi, J., Suto, A., Saitoh, Y., and Tsutsumi, K.-i. (2006). Binding of AIF-C, an Orc1-Binding Transcriptional Regulator, Enhances Replicator Activity of the Rat Aldolase B Origin. *Mol Cell Biol* 26, 8770-8780.
- Mohn, F., Weber, M., Schübeler, D., and Roloff, T.-C. (2009). Methylated DNA Immunoprecipitation (MeDIP). In, pp. 55-64.

- Montoya-Burgos, J.I., Boursot, P., and Galtier, N. (2003). Recombination explains isochores in mammalian genomes. *Trends in Genetics* 19, 128-130.
- Mugal, C.F., Wolf, J.B., von Grunberg, H.H., and Ellegren, H. (2010). Conservation of neutral substitution rate and substitutional asymmetries in mammalian genes. *Genome Biol Evol* 2, 19-28.
- Nagasawa, K., Kitamura, K., Yasui, A., Nimura, Y., Ikeda, K., Hirai, M., Matsukage, A., and Nakanishi, M. (2000). Identification and characterization of human DNA polymerase beta 2, a DNA polymerase beta -related enzyme. *J Biol Chem* 275, 31233-31238.
- Natale, D.A., Li, C.-J., Sun, W.-H., and DePamphilis, M.L. (2000). Selective instability of Orc1 protein accounts for the absence of functional origin recognition complexes during the M-G1 transition in mammals. *EMBO J* 19, 2728-2738.
- Necsulea, A., Guillet, C., Cadoret, J.C., Prioleau, M.N., and Duret, L. (2009). The relationship between DNA replication and human genome organization. *Mol Biol Evol* 26, 729-741.
- Nick McElhinny, S.A., Havener, J.M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A., and Ramsden, D.A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* 19, 357-366.
- Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25, 2730-2731.
- Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.
- Ovcharenko, I., Boffelli, D., and Loots, G.G. (2004). eShadow: a tool for comparing closely related sequences. *Genome Res* 14, 1191-1198.
- Perry, J., and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. *Curr Biol* 9, 987-989.
- Pfeifer, G.P. (2006). Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* 301, 259-281.
- Picher, A.J., Garcia-Diaz, M., Bebenek, K., Pedersen, L.C., Kunkel, T.A., and Blanco, L. (2006). Promiscuous mismatch extension by human DNA polymerase lambda. *Nucleic Acids Res* 34, 3259-3266.
- Polak, P., and Arndt, P.F. (2009). Long-range bidirectional strand asymmetries originate at CpG islands in the human genome. *Genome Biol Evol* 1, 189-197.
- Polak, P., Querfurth, R., and Arndt, P.F. (2010). The evolution of transcription-associated biases of mutations across vertebrates. *BMC Evol Biol* 10, 187.
- Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al. (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2, e168.
- Ponger, L., Duret, L., and Mouchiroud, D. (2001). Determinants of CpG islands: expression in early embryo and isochores structure. *Genome Res* 11, 1854-1860.
- Ponger, L.c., and Mouchiroud, D. (2002). CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics* 18, 631-633.

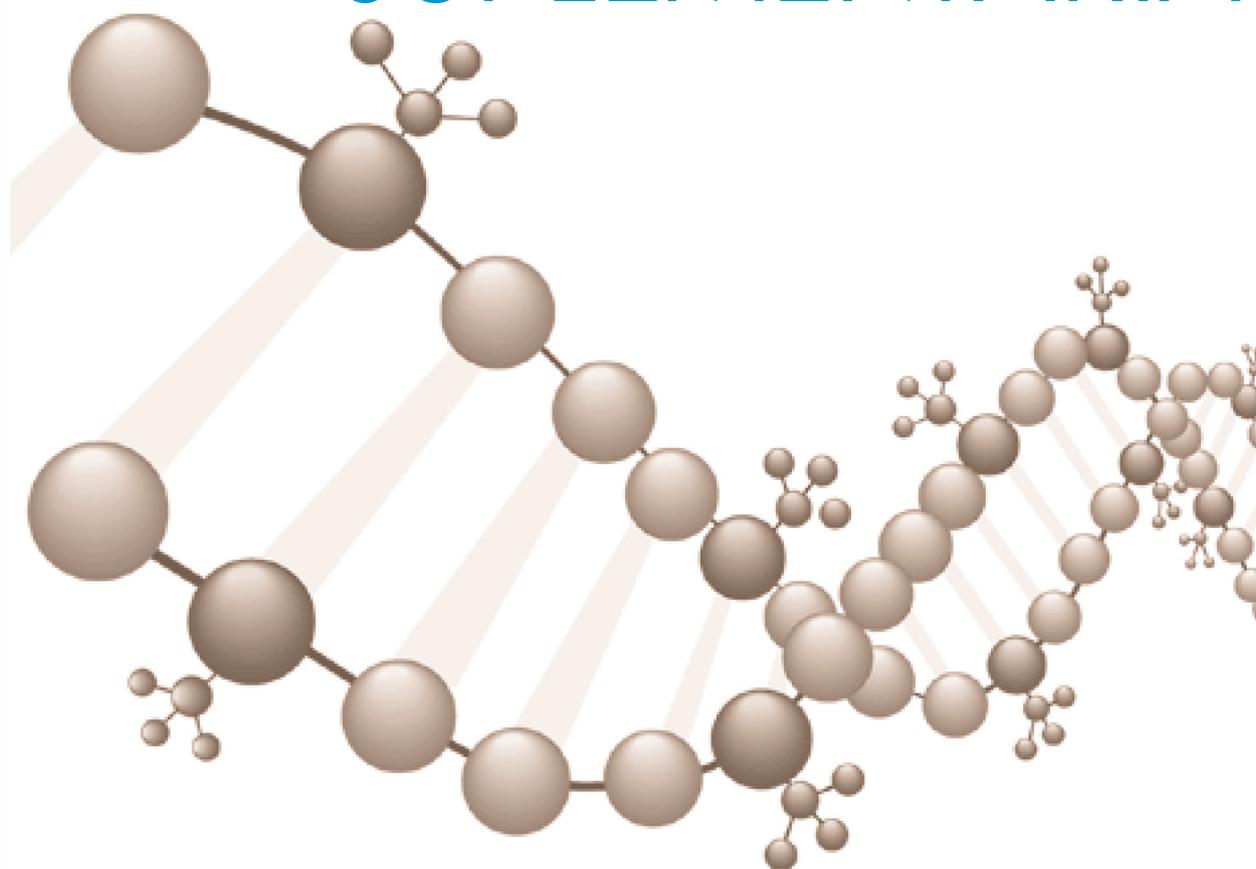
- Quaderi, N.A., Schweiger, S., Gaudenz, K., Franco, B., Rugarli, E.I., Berger, W., Feldman, G.J., Volta, M., Andolfi, G., Gilgenkrantz, S., et al. (1997). Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat Genet* 17, 285-291.
- Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M., et al. (2008). An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* 18, 1518-1529.
- Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* 138, 114-128.
- Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2009). A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* 106, 671-678.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425-432.
- Rein, T., Zorbas, H., and DePamphilis, M.L. (1997). Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol Cell Biol* 17, 416-426.
- Reynaud, C., Bruno, C., Boullanger, P., Grange, J., Barbesti, S., and Niveleau, A. (1992). Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. *Cancer Lett* 61, 255-262.
- Rideout, W.M., 3rd, Eversole-Cire, P., Spruck, C.H., 3rd, Hustad, C.M., Coetzee, G.A., Gonzales, F.A., and Jones, P.A. (1994). Progressive increases in the methylation status and heterochromatinization of the myoD CpG island during oncogenic transformation. *Mol Cell Biol* 14, 6143-6152.
- Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 19, 542-552.
- Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Péquignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes & Development* 12, 2108-2113.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* 20, 761-770.
- Saitoh, Y., Miyagi, S., Ariga, H., and Tsutsumi, K.i. (2002). Functional domains involved in the interaction between Orc1 and transcriptional repressor AOF that bind to an origin/promoter of the rat aldolase B gene. *Nucleic Acids Research* 30, 5205-5212.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory Press.
- Sasai, N., and Defossez, P.A. (2009). Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int J Dev Biol* 53, 323-334.
- Sasai, N., Nakao, M., and Defossez, P.-A. (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Research* 38, 5015-5022.
- Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103, 1412-1417.

- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., et al. (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 39, D38-D51.
- Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J* 23, 191-201.
- Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes & Development* 24, 1590-1595.
- Schweiger, S., Foerster, J., Lehmann, T., Suckow, V., Muller, Y.A., Walter, G., Davies, T., Porter, H., van Bokhoven, H., Lunt, P.W., et al. (1999). The Opitz syndrome gene product, MID1, associates with microtubules. *Proceedings of the National Academy of Sciences* 96, 2794-2799.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* 322, 1849-1851.
- Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* 5, e1000446.
- Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of the United States of America* 100, 15776-15781.
- Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15, 1034-1050.
- Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 283, 1164-1167.
- Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res* 14, 6711-6719.
- Spencer, C.C., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., Donnelly, P., Bentley, D., and McVean, G. (2006). The influence of recombination on human genetic diversity. *PLoS Genet* 2, e148.
- Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H. (2009). Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* 16, 564-571.
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 279, 27816-27823.
- Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9, 465-476.
- Suzuki, Y., Yamashita, R., Nakai, K., and Sugano, S. (2002). DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. *Nucleic Acids Research* 30, 328-331.

- Sved, J., and Bird, A. (1990). The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci U S A* 87, 4692-4696.
- Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., et al. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat Genet* 41, 572-578.
- Taira, T., Iguchi-Arigo, S.M., and Arigo, H. (1994). A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol Cell Biol* 14, 6386-6397.
- Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3740-3745.
- Takeshima, H., Yamashita, S., Shimazu, T., Niwa, T., and Ushijima, T. (2009). The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res* 19, 1974-1982.
- Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell* 60, 909-920.
- Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082-1086.
- Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Ofillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res* 14, 62-66.
- Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 17, 1469-1475.
- Uchiyama, Y., Takeuchi, R., Kodera, H., and Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie* 91, 165-170.
- Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev* 17, 1894-1908.
- Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20, 116-117.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.
- Weaver, J.R., Susiarjo, M., and Bartolomei, M.S. (2009). Imprinting and epigenetic changes in the early embryo. *Mamm Genome* 20, 532-543.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853-862.
- Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39, 457-466.
- Webster, M.T., and Smith, N.G. (2004). Fixation biases affecting human SNPs. *Trends Genet* 20, 122-126.

- Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M., et al. (2001). Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res* 11, 389-404.
- Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* 5, 695-705.
- Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y., and Ito, T. (2004). A Comprehensive Analysis of Allelic Methylation Status of CpG Islands on Human Chromosome 21q. *Genome Research* 14, 247-266.
- Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13, 335-340.
- Yoshikawa, K., and Aizawa, T. (1988). Enkephalin precursor gene expression in postmeiotic germ cells. *Biochemical and Biophysical Research Communications* 151, 664-671.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916-919.
- Zhao, Z., and Han, L. (2009). CpG islands: Algorithms and applications in methylation studies. *Biochemical and Biophysical Research Communications* 382, 643-645.
- Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. *Trends Genet* 24, 481-484.

INFORMACIÓN SUPLEMENTARIA



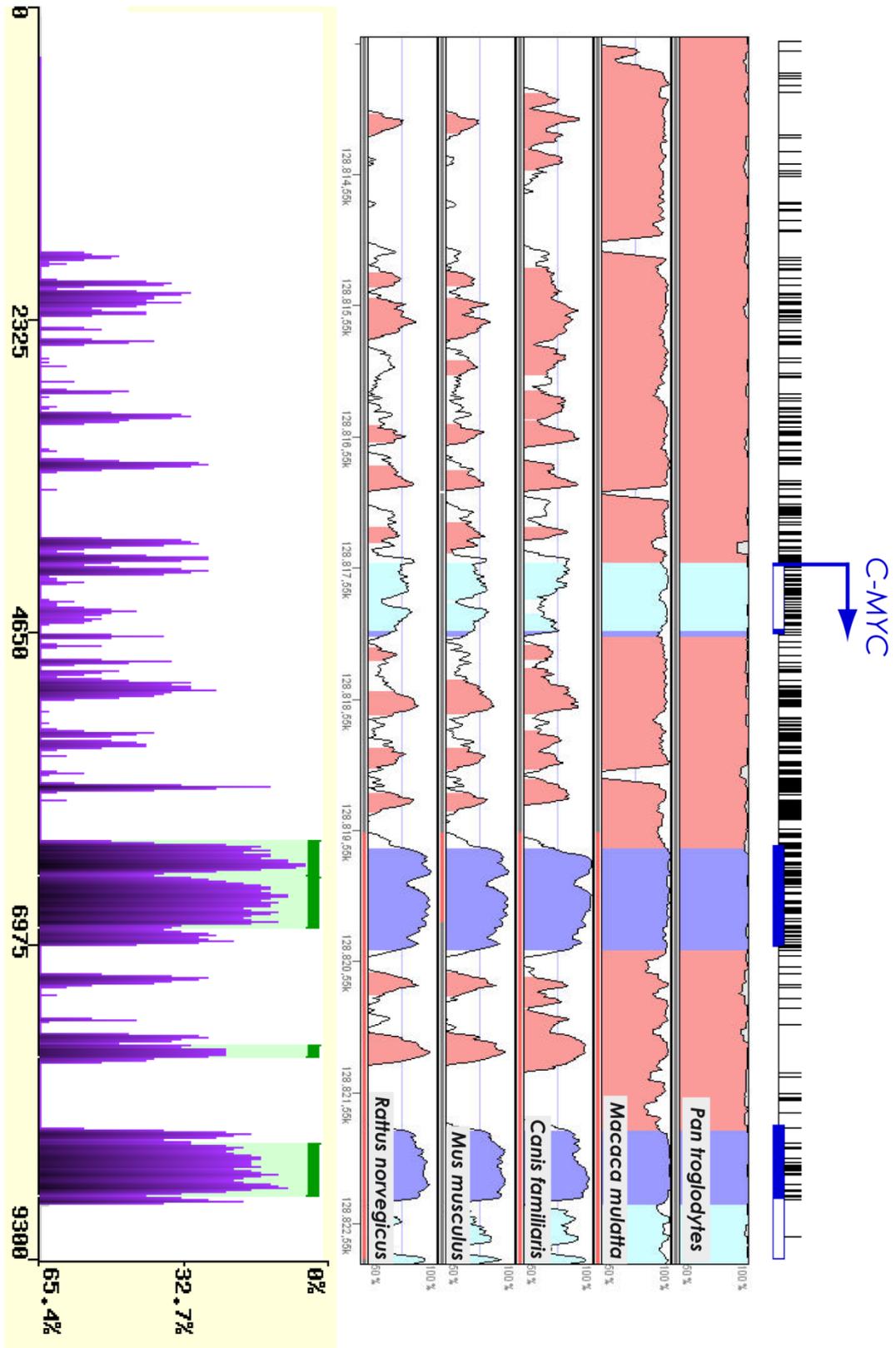


Figura S1. A. Análisis de la conservación filogenética del ORI asociado al gen C-MYC.

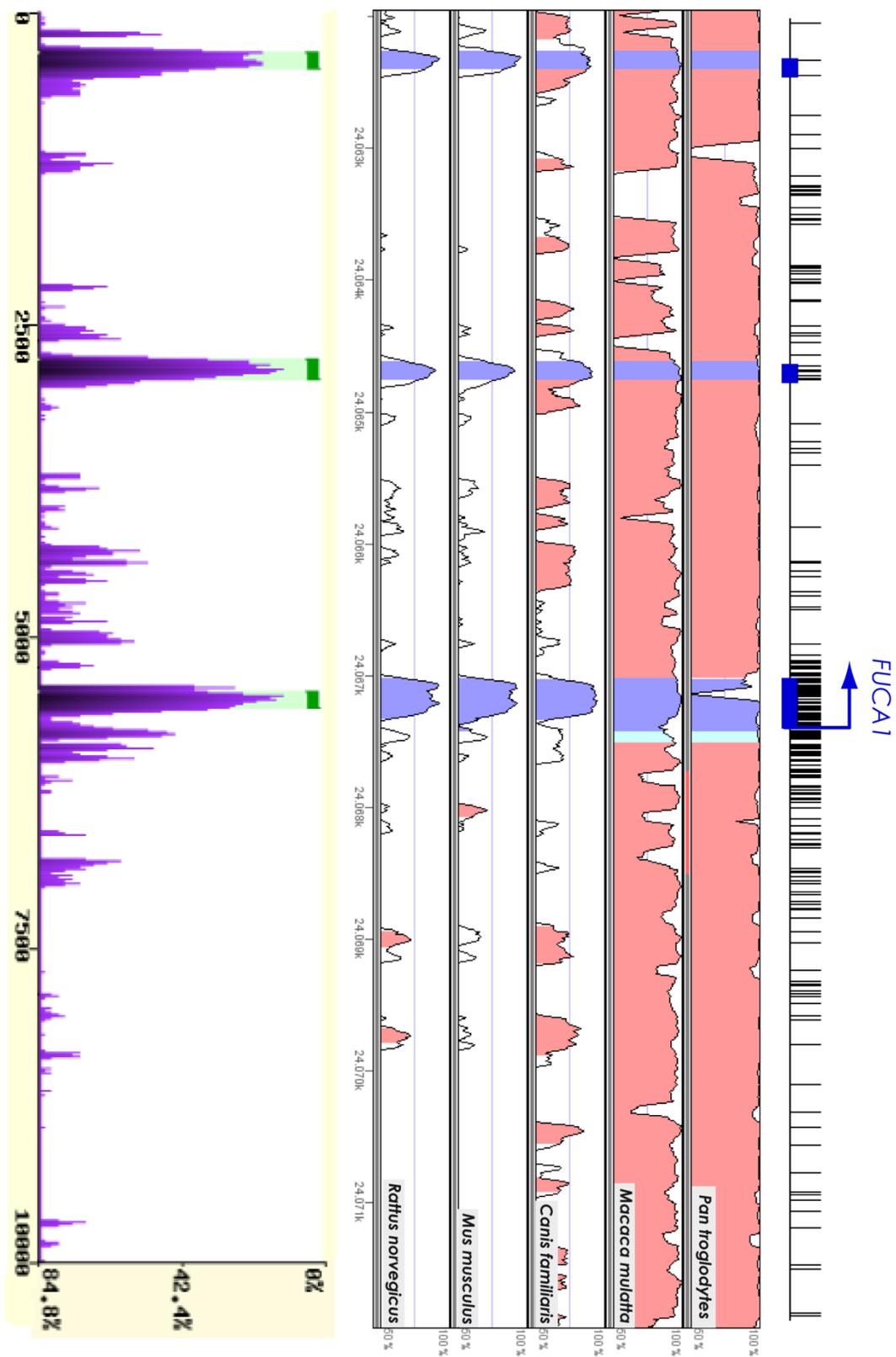


Figura S1. B. Análisis de la conservación filogenética del ORI asociado al gen *FUCA1*.

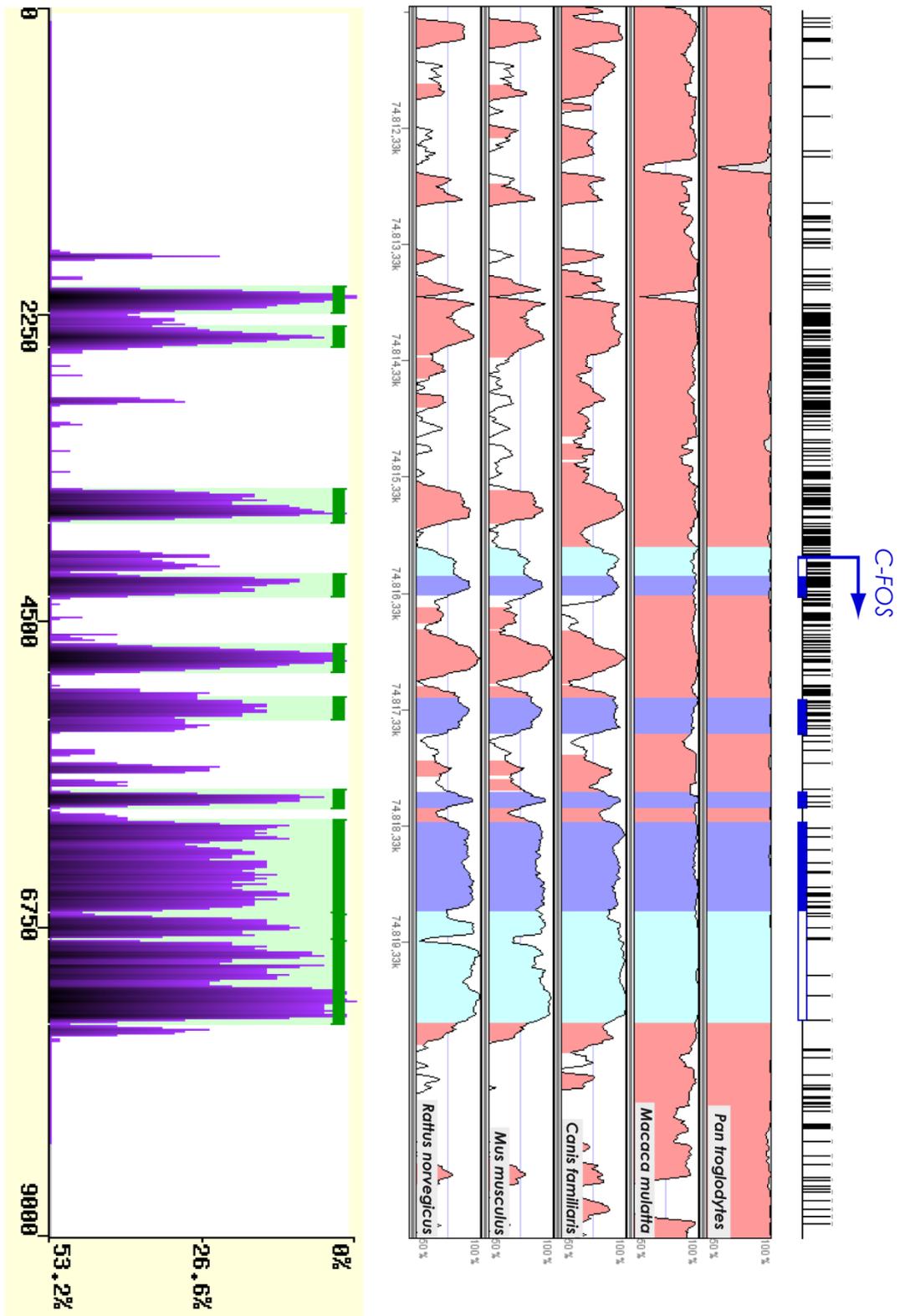


Figura S1. C. Análisis de la conservación filogenética del ORI asociado al gen C-FOS.

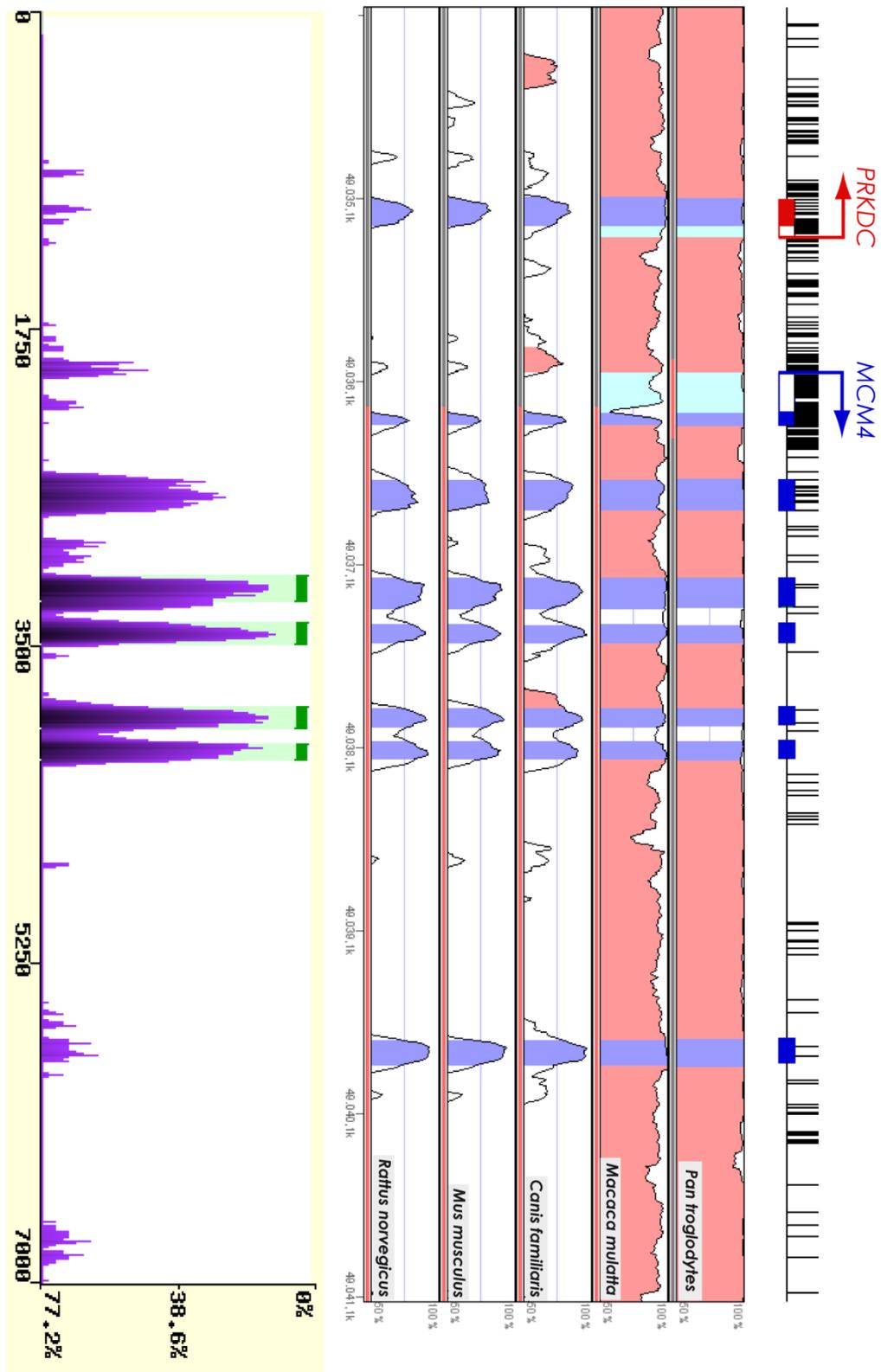


Figura S1. D. Análisis de la conservación filogenética del ORI asociado al gen MCM4.

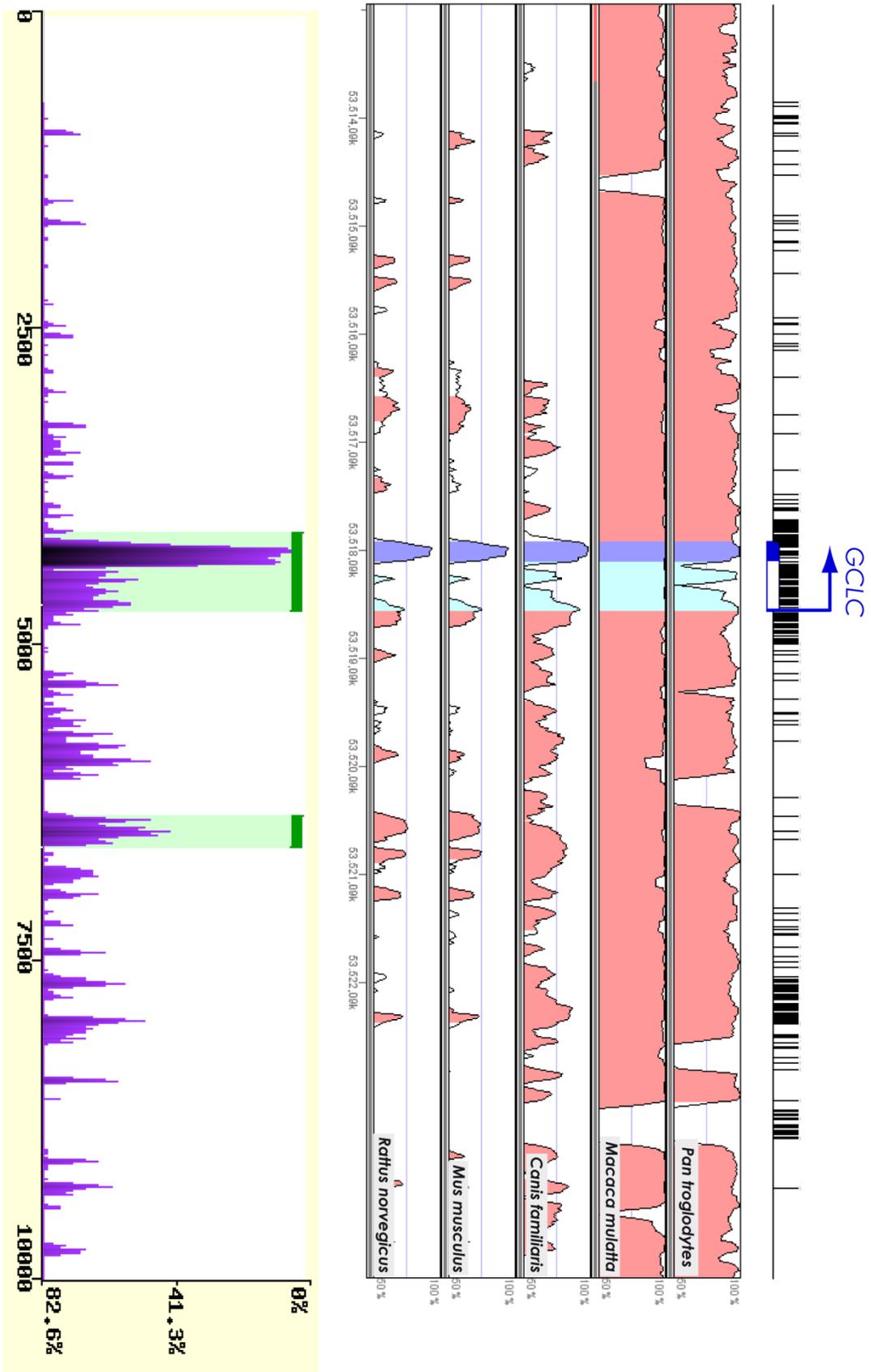


Figura S1. E. Análisis de la conservación filogenética del ORI asociado al gen GCLC.

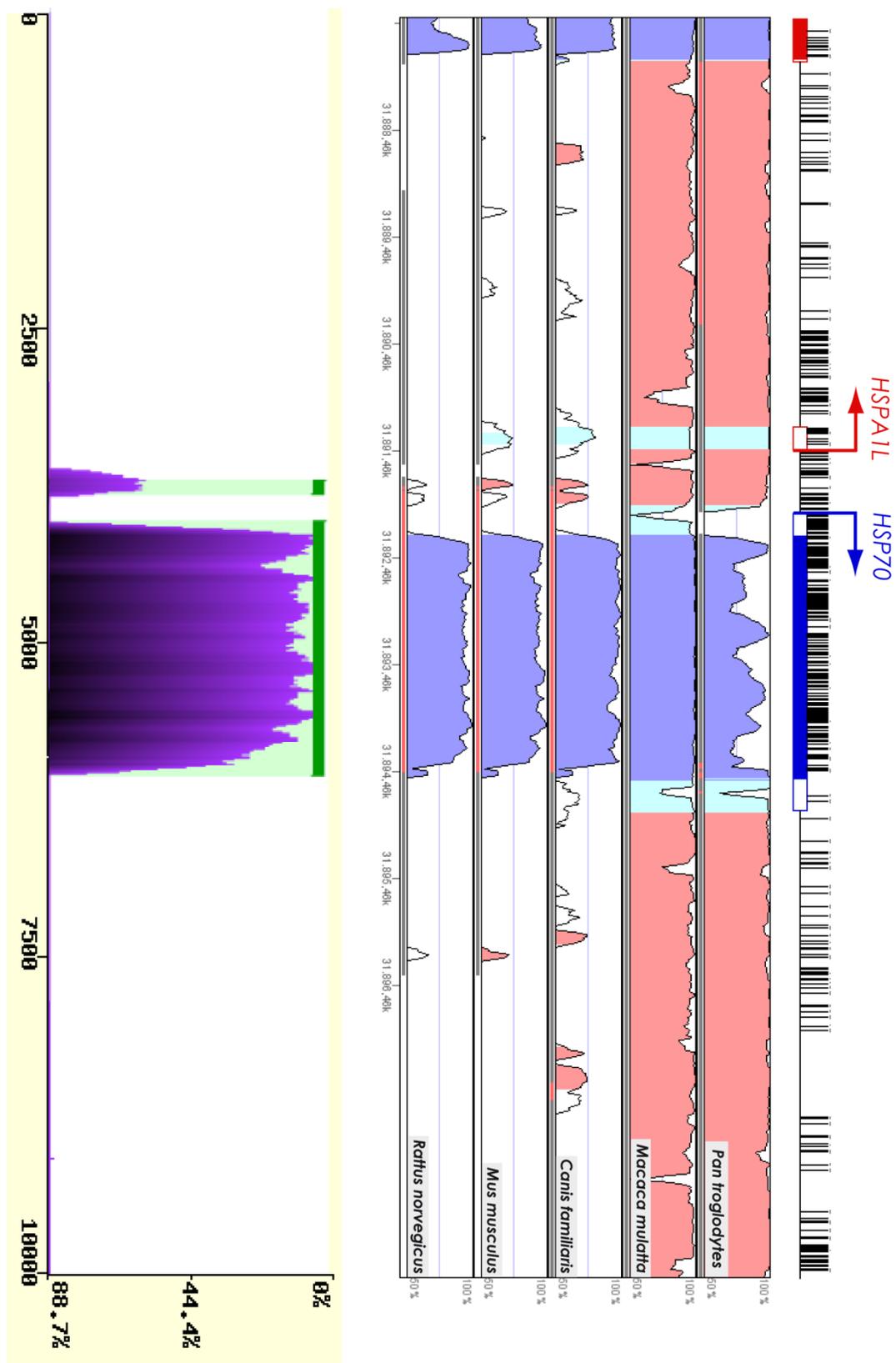


Figura S1. F. Análisis de la conservación filogenética del ORI asociado al gen HSP70.

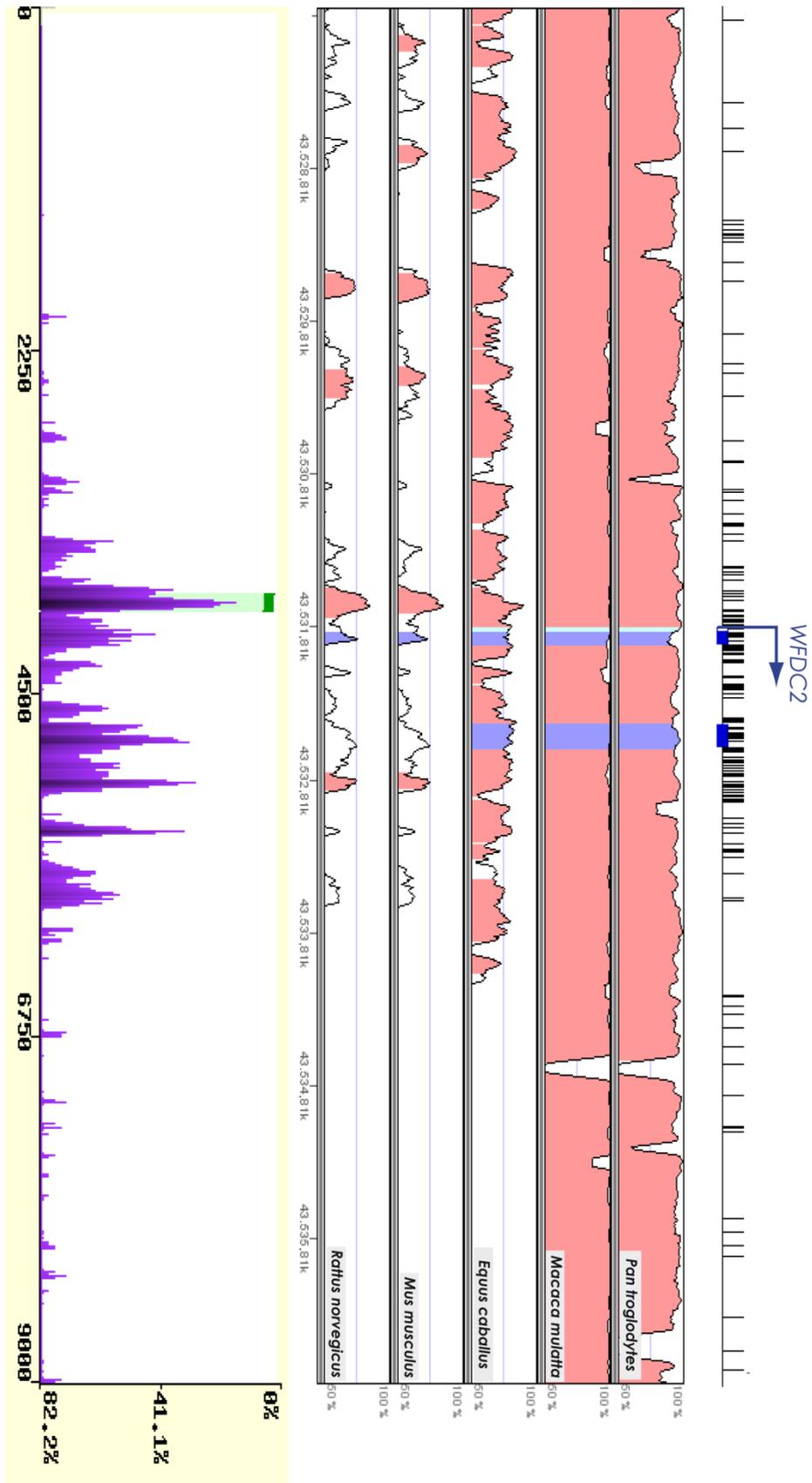


Figura S2. A. Análisis de conservación filogenética en el extremo 5' de la isla CpG asociada al gen WFDC2.

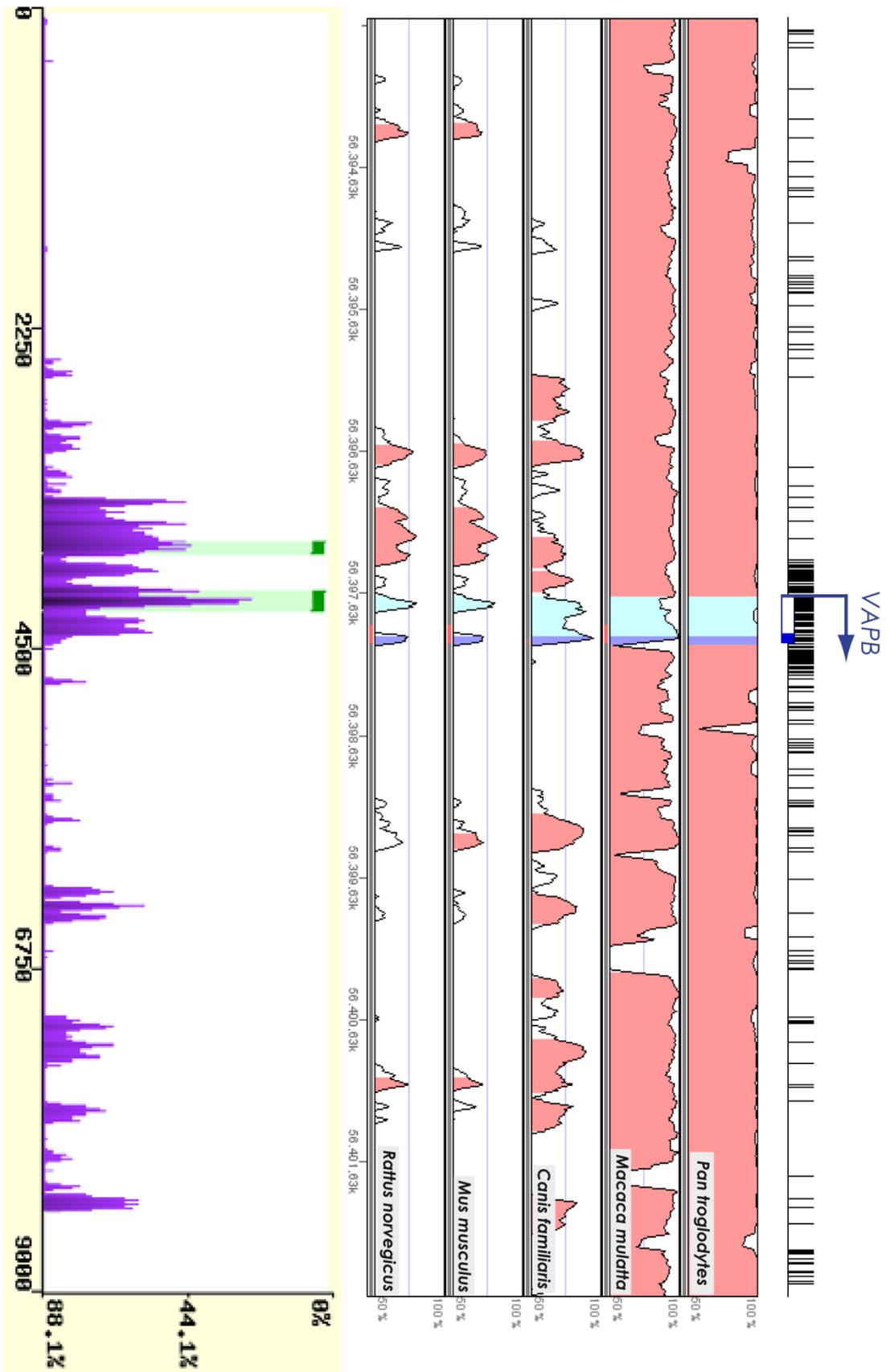


Figura S2. B. Análisis de conservación filogenética en el extremo 5' de la isla CpG asociada al gen VAPB.

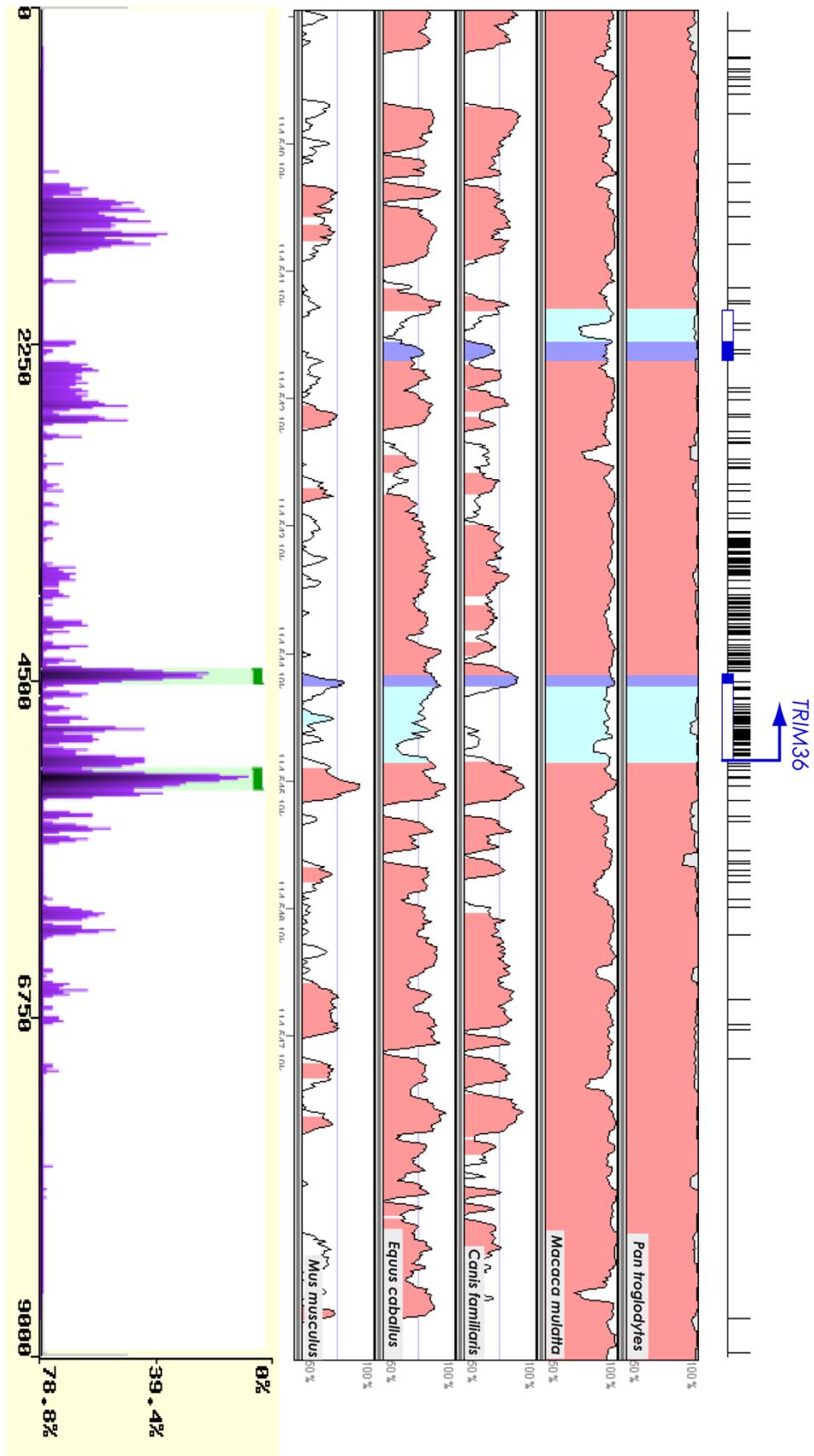


Figura S2. C. Análisis de conservación filogenética en el extremo 5' de la isla CpG asociada al gen *TRIM36*.

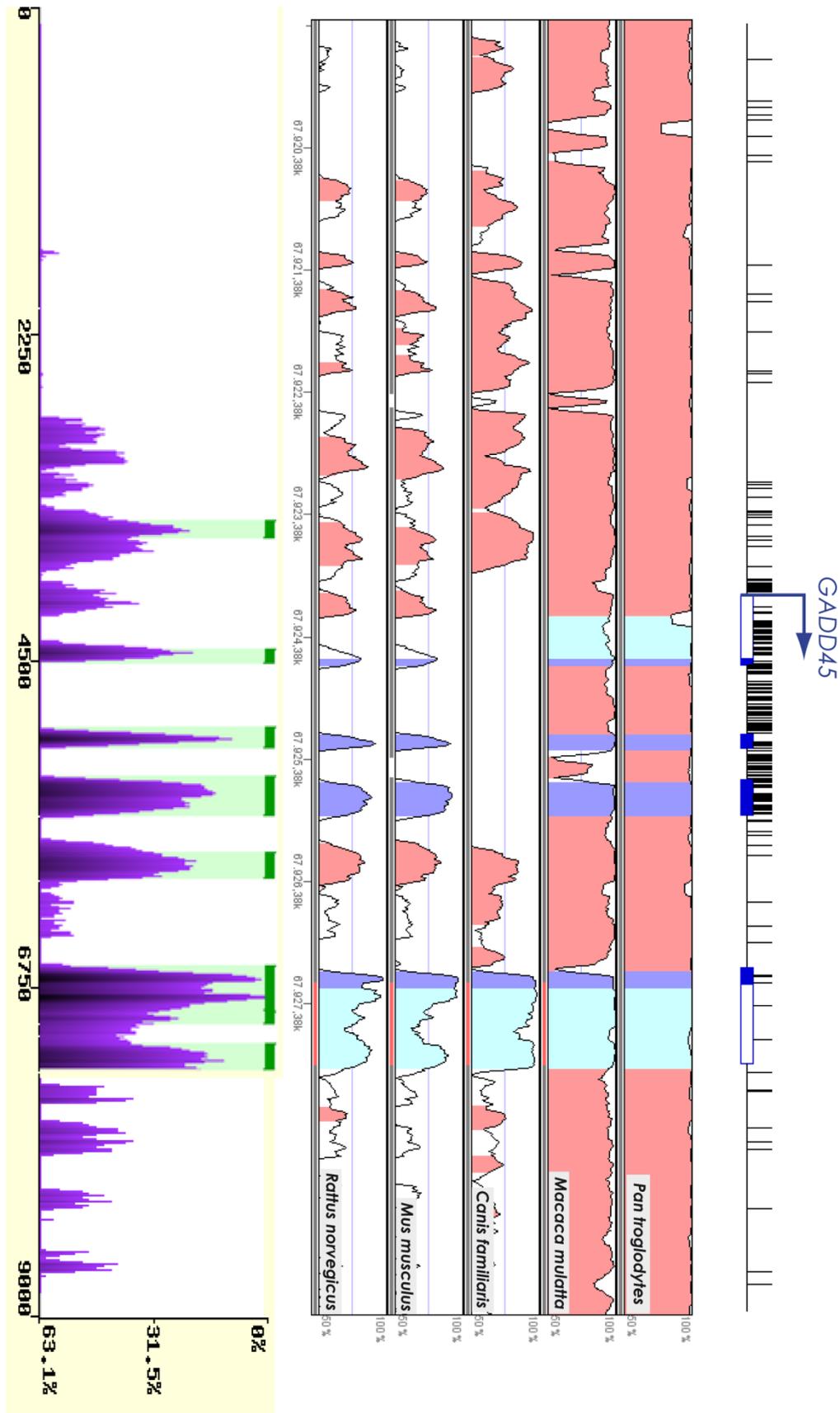


Figura S2. D. Análisis de conservación filogenética en el extremo 5' de la isla CpG asociada al gen GADD45A.

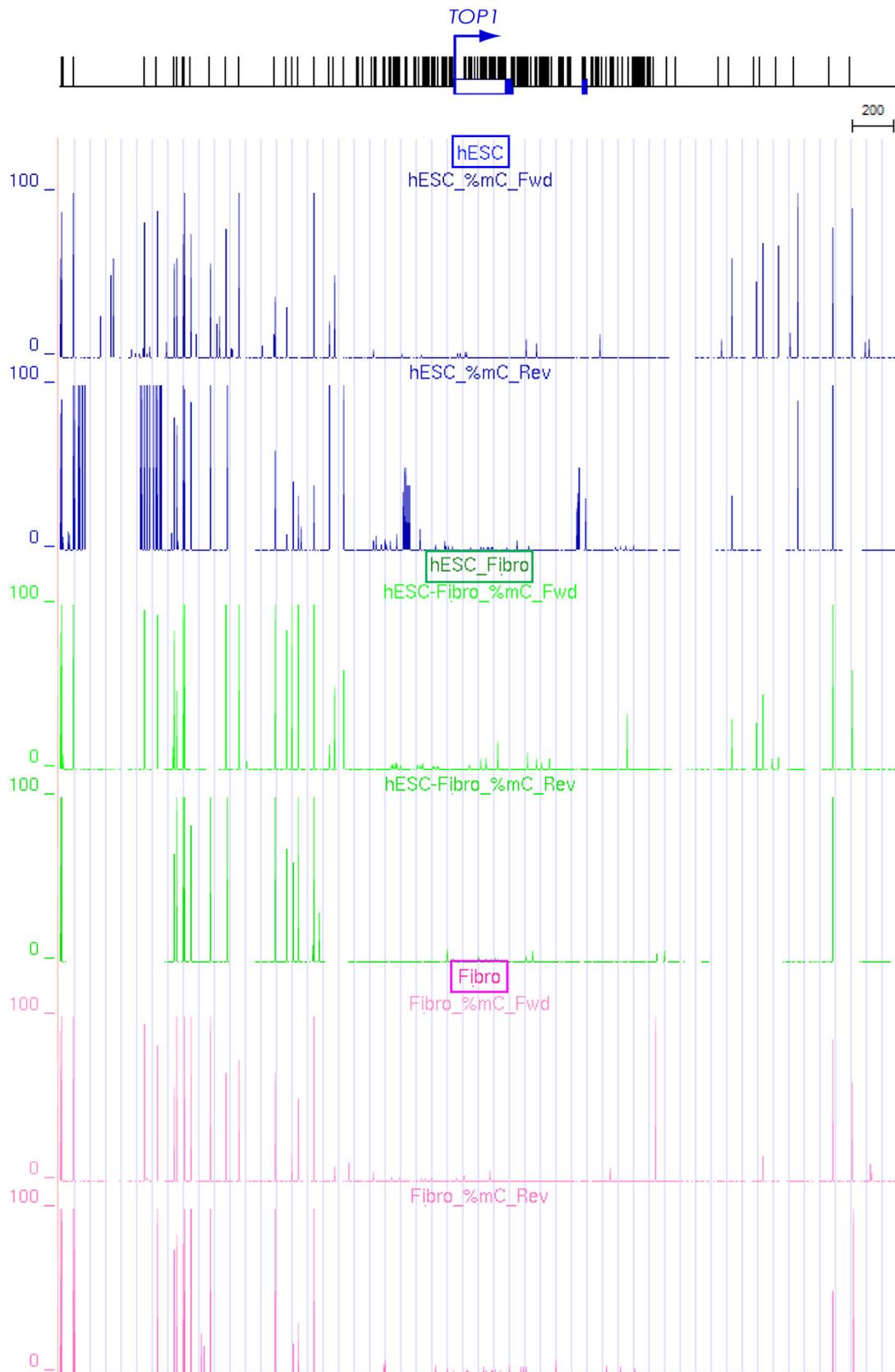


Figura S3. A. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para el gen *TOP1*.

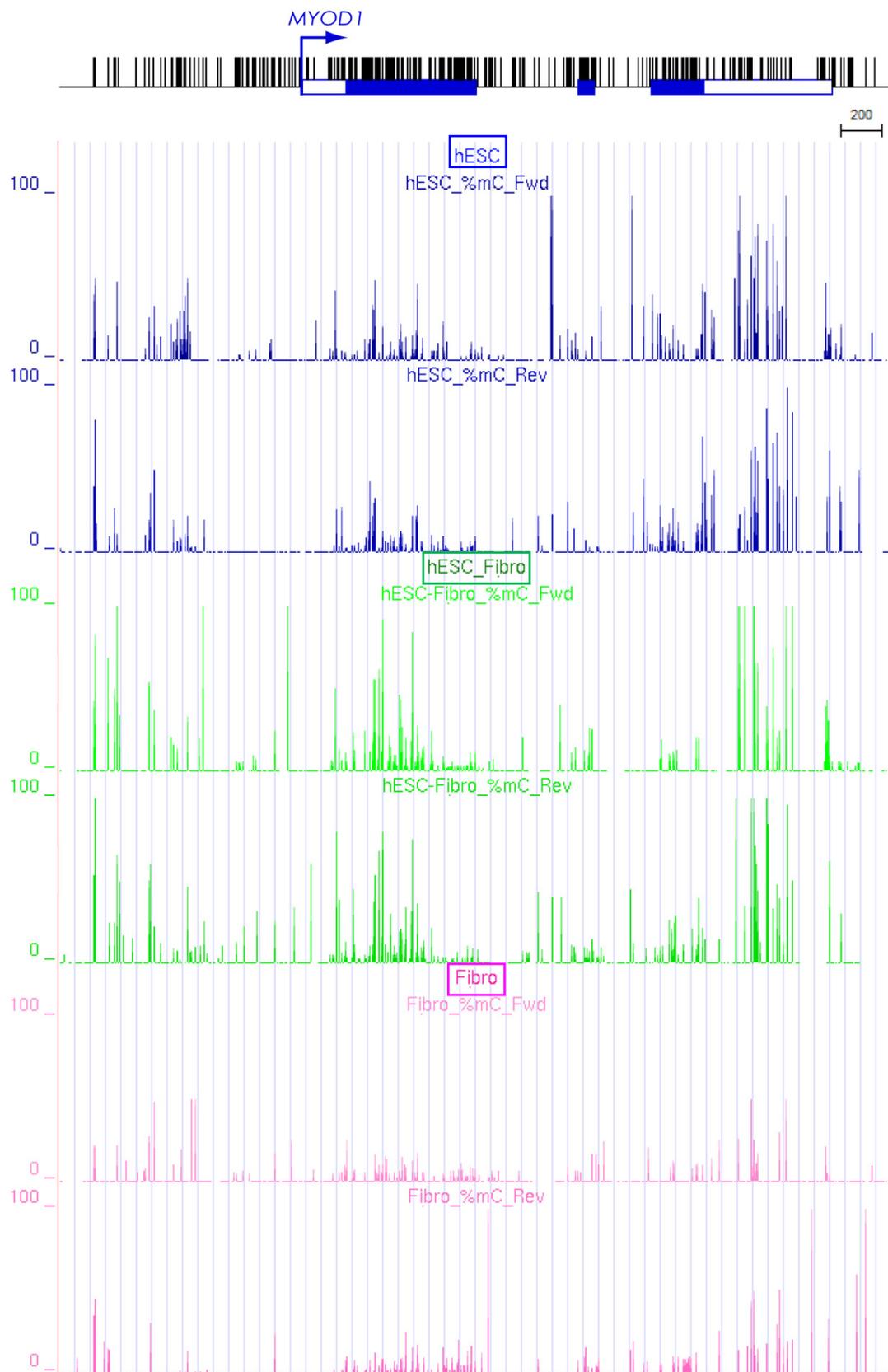


Figura S3. B. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para el gen *MYOD1*.

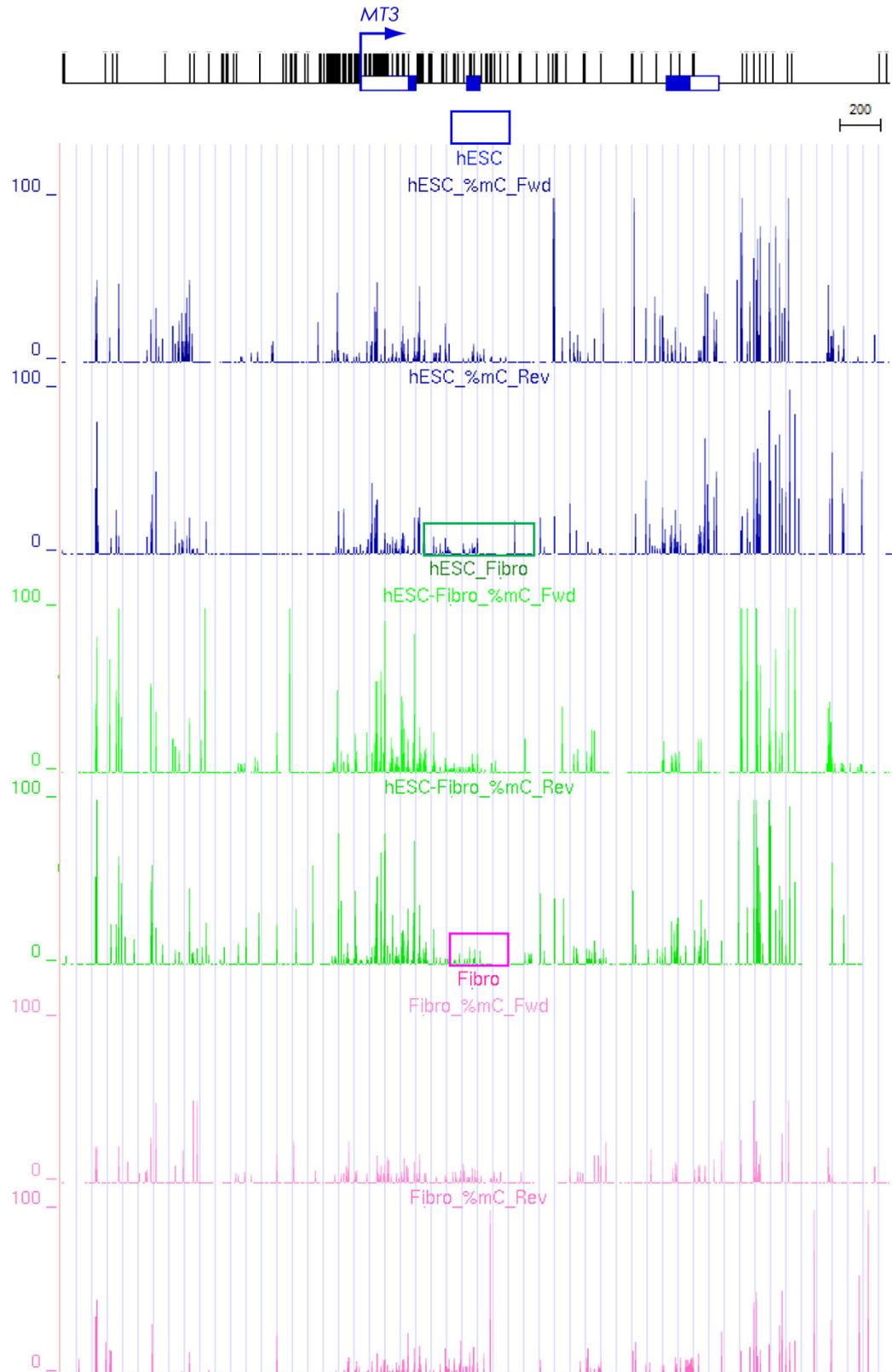
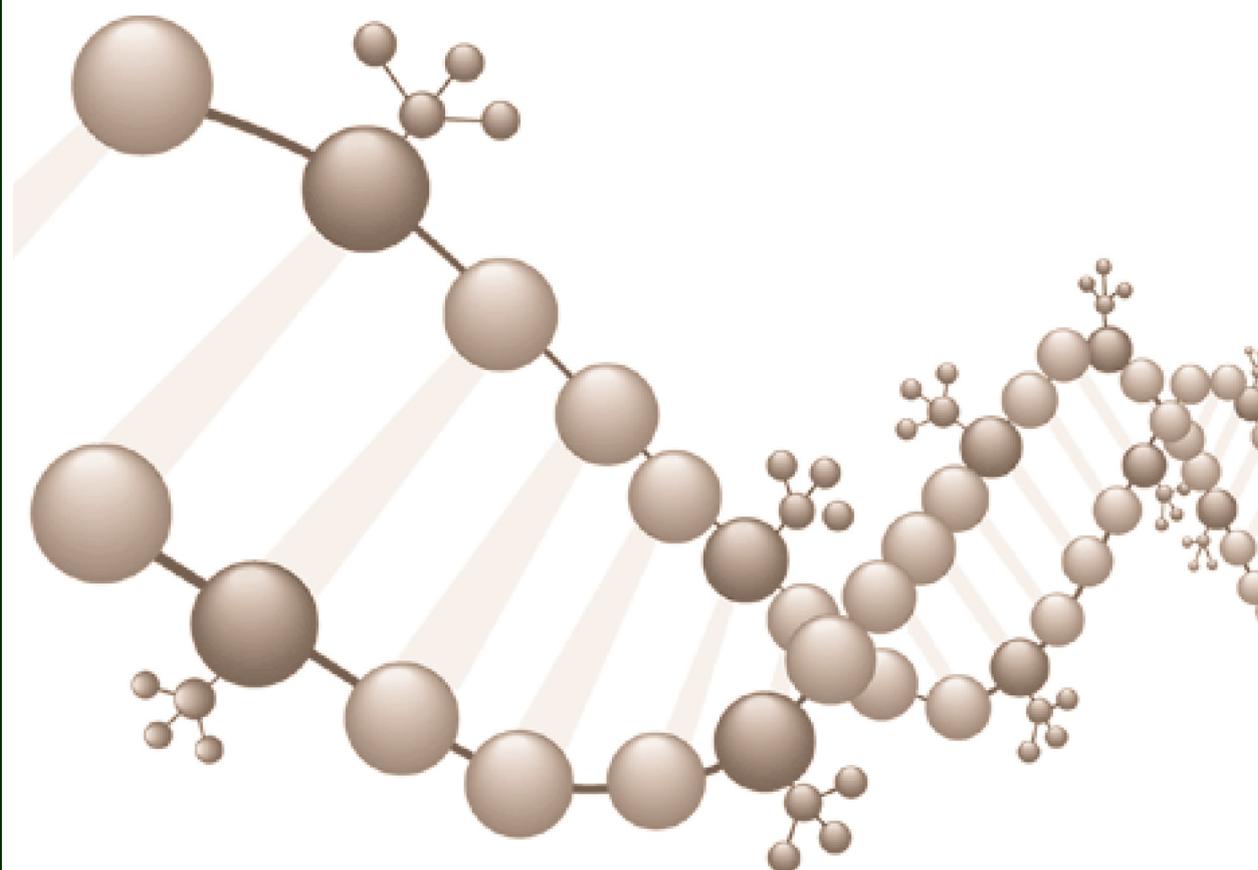


Figura S3. C. Detalle de los metilomas de hESC, fibroblastos derivados de hESC (hESC-Fibro) y fibroblastos de neonato (Fibro) para el gen *MT3*.

ABREVIATURAS



- A** - Adenina
- Asp** – Ácido aspártico
- BER** – Reparación por escisión de base – *Base Excision Repair*
- C** - Citosina
- °C** - Centígrado
- CAGE** – *CAP Analysis of Gene Expression*
- ChIP** - Inmunoprecipitación de cromatina - *Chromatin immunoprecipitation*
- CHO** – Ovario de hámster chino - *Chinese Hamster Ovary*
- DMEM** - *Dulbecco's Modified Eagle Medium*
- DNA** - Ácido desoxirribonucleico
- DNMTs** – DNA Metiltransferasas
- EDTA** - Ácido etilen diamino tetra acético - *Ethylene diamine tetraacetic acid*
- ES** – *Embryonic Stem*
- G** - Guanina
- Gly** - Glicina
- hESC** – Células madre embrionarias humanas – *Human Embryonic Stem Cells*
- ICR** – Centros reguladores de *imprinting* – *Imprinting Control Regions*
- Kb** - Kilobase
- µg** - Microgramo
- µl** - Microlitro
- µM** - micromolar
- M** - Molar
- Mb** - Megabase
- MBDs** – Proteínas de unión a DNA metilado – *Methyl Binding Proteins*
- MeDIP** – Inmunoprecipitación de DNA metilado – *Methylated DNA Immunoprecipitation*
- MeDIP-CHIP** – MeDIP combinado con la hibridación de *microarrays*
- mg** - Miligramo
- min** - Minutos
- ml** - Mililitro
- mM** - Milimolar
- ng** - Nanogramo
- NHEJ** - Unión de extremos no homólogos - *Non homologous end joining*
- nM** - nanomolar
- ORC** - Complejo de reconocimiento del origen - *Origin recognition complex*
- ORI** - Origen de replicación
- pb** - Pares de bases

ABREVIATURAS

PBS – Tampón fosfato salino - *Phosphate buffer saline*

PCR - Reacción en cadena de la polimerasa - *Polymerase Chain Reaction*

PGC – Células precursoras germinales – *Primordial Germ Cells*

pre-RC - Complejo prerreplicativo - *Pre-replicative complex*

qPCR – PCR cuantitativa

RNA - Ácido ribonucleico

RNApol II – RNA polimerasa II

rpm - Revoluciones por minuto

SDS – Dodecilsulfato sódico

T – Timina

TE - Tris/EDTA

tRNA – RNA de transferencia

TSS – Sitio de inicio de la transcripción – *Transcription Start Site*

U – Unidades

Val - Valina