



DEPARTAMENTO DE BIOLOGÍA MOLECULAR

Interrelación entre la estructura de la cromatina, la transcripción génica y la iniciación de la replicación del DNA en células de mamífero y en *Leishmania major*

RODRIGO LOMBRAÑA PASCUAL

Madrid, Enero 2017

Fotografía de portada: Análisis de fibras de DNA de *Leishmania major* mediante la aplicación sucesiva de análogos de nucleótidos fluorescentes. Se pueden observar fibras con varios orígenes de replicación así como los maxicírculos de DNA característicos de estos tripanosomátidos.

Micrografía: Ricardo Almeida y Alba Álvarez

Montaje: Eduardo Lombraña-Pascual



**DEPARTAMENTO DE BIOLOGÍA MOLECULAR
FACULTAD DE CIENCIAS
UNIVERSIDAD AUTÓNOMA DE MADRID**

**Interrelación entre la estructura de la cromatina, la
transcripción génica y la iniciación de la replicación del
DNA en células de mamífero y en *Leishmania major***

RODRIGO LOMBRAÑA PASCUAL

Ldo. en Biotecnología

Directora de Tesis: Dra. María Gómez Vicentefranqueira
Centro de Biología Molecular “Severo Ochoa” (UAM/CSIC)

Madrid, Enero 2017

Los trabajos que conforman la realización de esta Tesis Doctoral han sido financiados por el Ministerio de Ciencia e Innovación (actual Ministerio de Economía) mediante una ayuda de Formación de Personal Investigador (FPI) [referencia BES-2011-044318] dentro del proyecto BFU2010-18992

Al igual que las montañas parecen descansar sobre los paralelos raíles férreos, los nucleosomas “cubren” el DNA. La polimerasa intenta acceder entre los nucleosomas, como la nube del valle. Bajo ella se encuentran los orígenes...



“Paisaje nucleosomal”:

Fotografía: Rodrigo Lombraña-Pascual y Julieta Arancibia Llana

Montaje: Eduardo Lombraña-Pascual

A **Feli, Mena** y los que ya no están,

por ser la voz de la experiencia

A mis padres, **Eduardo y Ludi**, a **Edu** y a **Julieta**,

por ser pasado, presente y futuro

A **Gabriela**

por ser el porvenir

“La ciencia es una forma de pensar, mucho más que un conjunto de conocimientos. [...] No es perfecta, con frecuencia se utiliza mal. No es más que una herramienta, pero es la mejor herramienta que tenemos: se corrige a sí misma, está siempre evolucionando y se puede aplicar a todo.”

“La imaginación frecuentemente nos llevará a mundos que jamás fueron. Pero sin ella, no iremos a ningún lado”

Carl Sagan

En primer lugar me gustaría agradecer a María Gómez la oportunidad que me ha brindado para introducirme durante todos estos años en el campo de la investigación científica. Ya desde los tiempos de Salamanca, has sabido motivar y estimularme, intentando que sacase lo mejor de mí cada día. Decía JD Watson que es importante elegir un supervisor joven que esté en su apogeo de productividad intelectual; no sé si esta etapa habrá sido el máximo pero desde luego, has conseguido transmitirme gran cantidad de valores científicos y personales que estoy seguro de que me serán tremendamente útiles en el futuro. Gracias por tu entusiasmo.

Gracias también a los miembros del grupo, a los “fijos” Ricardo, Josemi y Sofía (lo siento, siempre seguirás siendo del 405...) y a los que habéis pasado (y hecho pasar) tan buenos ratos: Gonzalo, Alba y Cristina. Os agradezco todo el apoyo que siempre me habéis ofrecido tanto científicamente como en los ratos de ocio. Sinceramente, creo que es esa mezcla tan bien homogeneizada la que nos caracteriza y, aunque esté mal decirlo, la que probablemente nos convierte en un gran ejemplo de equipo consiguiendo hacer llevaderos hasta los peores momentos. Porque sois mucho más que compañeros, grandes amigos. ¡¡Gracias equipo!! ¡*True Survivor!*! No me quiero olvidar de aquellos que habéis pasado breves estancias entre nosotros: Dani, Kenza, Nicoló, David, etc. dejando vuestra huella imborrable y grandes recuerdos.

Gracias también al grupo de Crisanto Gutiérrez por las discusiones, sin duda interesantes, surgidas en los seminarios conjuntos y en el día a día. También por supuesto, en todos los ámbitos que acabamos compartiendo a partir de este trabajo: Béné, Sofía, María, Carla, Nuria, etc. También a las nuevas incorporaciones Clara y Aitor, ¡os deseo un futuro brillante! A Joana, que no sé cómo catalogarte: mentora y amiga desde Salamanca, un gran apoyo y sobretodo, uno de mis referentes más importante; te deseo lo mejor para el futuro, porque ya sabes que todos queremos ser como tú. A Zaida, gran compañera de penas y alegrías, mi nueva hermana, con quien las jornadas eternas se hacen cortas; esa mezcla especial (e inusual) de gran persona y mejor científica (o viceversa); sabes que parte de esta Tesis es tuya, porque no hubiese sido lo mismo sin conocerte en aquel Far West(ern).

Gracias a José María Requena, por su constante entusiasmo y empeño en que la colaboración llegase a un punto fructífero. Gracias a todos los miembros de su grupo con los que espero seguir compartiendo buenos momentos. Un recuerdo especial a César, por todo lo que he aprendido de ti durante las charlas intentando domar al parásito; también por tu constante disposición a ayudarme en estos años de proyecto conjunto.

A Luis Blanco, tutor de esta Tesis pero sobre todo, un referente como científico de excelencia. Gracias también por el interés mostrado hacia mis avances cada vez que hemos coincidido, ya sea en un seminario o en una conversación informal de pasillo; tus ideas científicas siempre resultan de gran utilidad y tu entusiasmo es algo difícil de ignorar.

Quiero agradecer la disposición de David Pisano a que alguien ajeno a la Unidad de Bioinformática del CNIO pudiera dar sus primeros pasos en el análisis computacional de datos. También a los demás miembros de la Unidad que me mostraron un universo nuevo para mí y me acogieron con cariño durante aquella ya lejana semana. Especialmente gracias a Osvaldo Graña por su paciencia pedagógica durante esos días y los que nuestros trabajos se han cruzado *a posteriori*, por tu entusiasmo y entrega desde lo profesional a lo musical. Gracias también a Miriam Rubio por su paciencia y disposición para aclarar dudas sobre el funcionamiento de RUBioSeq.

Trabajar en un centro como el CBM es un privilegio y motivo de orgullo. Pero además, también te permite contar con excelentes profesionales que son grandes personas. Gracias a Ugo, por colaborar en estos trabajos así como por las estimulantes charlas mantenidas; siempre se agradece poder contar con un punto de vista matemático de lo que nos rodea. Igualmente a Ramón, David y Alfonso por mantenerse colaborativos y pacientes con los que os asaltamos con dudas variopintas (y repetitivas) y aun así, facilitar siempre soluciones. También agradezco la predisposición a ayudar de Alberto Rastrojo en aquellos inciertos primeros pasos en el campo bioinformático. Gracias al Servicio de Genómica por vuestra constante disposición e interés por los experimentos, con un recuerdo especial para Alberto, al que me unen las raíces y a quien deseo lo mejor para su futuro. Gracias en general a todo el personal del Centro, por facilitarnos el trabajo, muchas veces de manera anónima y discreta. También gracias a esos investigadores con los que un cruce en el pasillo o una petición de enzima puede desembocar en una conversación que va desde insuflar ánimos a proponerte una idea significativamente interesante. Me vienen a la cabeza especialmente Laura Mojardín, Ignacio, Kiki, José Antonio, César, Elena... También incluso de fuera del CBM como Juan y Karolina o Arturo y Fábila -a vosotros además, gracias por vuestras enseñanzas e intentivas con las técnicas de 2D-. No me olvido de nuestros primeros compañeros de laboratorio y que siempre, al cruzarnos por el Centro, siguen manteniendo ese interés por nuestro trabajo: Pepa, Patricia y Ángel.

Lo bonito de la multidisciplinariedad es que se refleja las amistades con las que he podido compartir tan buenos ratos a lo largo de estos años. Desde los grupos de moscas -Sandra, Mer, Antonio, etc.- hasta los *cruzies* -Néstor, Carlos, Fernando, Bea, Cristóbal, Maika, Fran, Alba, etc-. Muchísimas gracias por vuestra amistad, especialmente durante los primeros años en Madrid. Espero que os vaya genial allá donde os lleve la vida. También a Julien, mi otro gran referente como joven científico y, junto con David, gran amigo; con quienes comencé compartiendo piso y he acabado siendo parte de su familia. Gracias por vuestros constantes consejos, tanto en las visitas como desde la distancia.

Thanks a lot to Greg Kudla for giving me the opportunity to enjoy a great stay in Edinburgh. You are one of the most brilliant and talented scientist that I have had the opportunity to meet. Thanks also to the people in that lab, specially Olga and Miriam for your friendship. Thanks to Yvonne and her family (who was also mine!). Thank you for offering me a different -and authentic- experience in those great lands of Scotland, thanks for being my family there.

Dicen que soy un tipo aferrado a sus orígenes, es lógico porque la familia nos dice de dónde venimos y, probablemente, hacia dónde vamos. Además, habéis sido y sois un apoyo y ayuda fundamental y constante. Gracias a mis padres, Eduardo y Ludi, por la educación que me habéis brindado siempre. Me gustaría que esta Tesis sirva de compensación a tantos esfuerzos invertidos en mi educación y formación. Gracias también a Edu, por tu ayuda técnica pero también por esa picardía que siempre acaba haciendonos pasar un buen rato a los que te rodeamos. Gracias a Lara, por el constante apoyo y la ayuda idiomática. Es imposible nombrar a todos los primos, primas, tíos y tías pero gracias a todos y cada uno de vosotros por el interés que siempre mostráis así como por la constante motivación. Sin todos vosotros, esto sería aún más complicado. Un recuerdo especial a mis dos abuelas, Feli y Mena, por hacerme afortunado de tenerlas; ellas, que son el reflejo de que el progreso no puede hacer olvidar de dónde procedemos y de que la sabiduría es un tanto poliédrica.

A la otra familia, María Jesús, Javier, Meli, Pauli y ahora también, Gabriela. Por todo el apoyo desde que me conocisteis y, en especial, durante estos (largos) meses de escritura. Gracias extensivas a todos los demás, a "los de Villablino", también familia al fin y al cabo. Gracias a los Dacosta-Laiz, a Roberto y Pili, a Llamas, a todo el grupo de amigas, etc. porque siempre intentáis que tengamos lo mejor para poder enfrentarnos a nuestros propios desafíos.

Dicen que los amigos son la familia que se elige. Desde luego, no lo pude hacer mejor, porque tanto si nos vemos de año en año, como si pasamos juntos una temporada, el resultado siempre es fantástico. A todos los que os he conocido fruto de estos años en Madrid -Almu (gran redescubrimiento)-, y más recientemente (pero no por ello tarde) Boris, Paula, Alberto, etc.- tengo que sumar los ánimos que tantos me habéis seguido transmitiendo ya sea desde Guardo -Javi, Pablo, Adri, Rober, Rubén, Julito, Lidia, Sara, Ceci (a ti además, gracias por los buenos ratos y consejos en Escocia), etc.- como desde León. Gracias a las *sanchas* -David, Pablo, Sara, Javi, etc.-. porque desde aquellos años siempre habéis confiado en mis posibilidades y animado a seguir por este camino. A los *quintillos* -Iñaki, Diego, Marcos, Ángel, Álvaro, Pepe, María, etc- por la suerte que ha sido compartir tanto con personas tan brillantes como vosotros, capaces de convertir un congreso en una juerga y cualquier juerga en un congreso; porque surcar los mares, celebrar bodorrios o simplemente, tomar unas tapas sin un sabor de fondo a Bio(tecno)logía no sería lo mismo... Gracias a todos vosotros, amigos, porque para este trabajo, no sólo han sido necesarios los consejos científicos... y ahí habéis estado siempre.

A ti Julieta, ¿qué decirte? Si ya me habías demostrado que no es cierto que la paciencia tenga un límite, si me has servido de apoyo cuando parecía que caía, si cuando no había luz, tu la encendías... Hemos compartido ya tantas cosas, he llevado tanto el laboratorio a tu vida, que no concibo este trabajo sin ti a mi lado ("I'd be a candle without a wick, a Watson without a Crick"). El lenguaje tiene palabras para casi todo y son pocas las veces que realmente no alcanzan; esta es una de esas ocasiones. No puedo agradecerte todo simplemente con discursos, tómate esta Tesis como algo que también te pertenece.

Gracias a todos mis educadores, especialmente aquellos que habéis sido además maestros y, en general, a todos las personas que a lo largo de mi vida, habéis contribuido a que haya tomado las decisiones que culminan en este trabajo. Gracias por "soñar que ese navío llevará nuestra carga de palabras hacia puertos distantes, hacia islas lejanas".



ÍNDICE

Resumen/Summary	5
Lista de abreviaturas	9
Introducción	13
1. Replicación del DNA	15
1.1. Interés biológico y progreso histórico	15
1.2. Metabolismo de la iniciación de la replicación del DNA en eucariotas: licenciamiento y activación	18
1.3. Orígenes de replicación en diversos modelos: especiación	20
1.4. Análisis de la iniciación de la replicación en poblaciones celulares: eficiencia de activación de los orígenes de replicación	23
2. Cromatina y regulación genómica	25
2.1. Estructura de la cromatina y expresión génica	26
2.2. Estructura de la cromatina y replicación del DNA	27
Objetivos	31
Materiales y métodos	35
1. Manipulación de material biológico	37
1.1. Células de ratón	37
1.2. Células de <i>Leishmania major</i>	38
2. Purificación de DNA genómico de alto peso molecular	38
2.1. Células de ratón	38
2.2. Células de <i>L. major</i>	39
3. Purificación de intermediarios de replicación de pequeño tamaño (SNS)	39
3.1. Células de ratón	39
3.2. Células de <i>L. major</i>	42
4. Purificación de DNA mononucleosomal	42
4.1. Ensayo de escaneo de nucleosomas (NuSA)	44
5. PCR cuantitativa (qPCR)	44
6. Secuenciación masiva (NGS)	45
6.1. Preparación de librerías de SNS-seq y MN-seq de <i>L. major</i>	45
6.2. Secuenciación	46
7. Análisis bioinformáticos	46
7.1. Análisis de datos de NGS de repositorios públicos provenientes de células de ratón y humano	46
7.1.1. Selección de los sitios de inicio de la transcripción (TSS)	46
7.1.2. Utilización de datos accesibles de células de ratón	46

7.1.3. Utilización de datos accesibles de células humanas.....	47
7.1.4. Manejo y tratamiento de los datos.....	47
7.2. Análisis de datos de NGS de <i>L. major</i>	48
7.2.1. Anotación de características genómicas a partir de datos públicos..	48
7.2.2. Alineamiento de lecturas de NGS.....	49
7.2.3. Definición de regiones significativas y construcción de perfiles	50
7.2.4. Manejo y tratamiento de los datos.....	51
Resultados	53
Parte I	55
1. Estudio genómico de la estructura de la cromatina y la iniciación de la replicación en células de ratón.....	55
2. Análisis local de alta resolución de la estructura de la cromatina y la iniciación de la replicación en células de ratón.....	56
3. Estudio genómico de la estructura de la cromatina y la iniciación de la replicación en células humanas.....	63
Parte II	69
1. Anotación del genoma de <i>L. major</i>	69
2. Análisis genómico de la iniciación de la replicación en <i>L. major</i>	74
3. Análisis genómico de la organización nucleosomal en <i>L. major</i>	85
4. Relación entre la estructura de la cromatina y los mecanismos de replicación y transcripción en <i>L. major</i>	90
Discusión	97
1. Sesgos y motivos en la secuencia del DNA que podrían favorecer la iniciación de la replicación.....	101
2. El paisaje nucleosomal refleja la probabilidad de iniciación de la replicación del DNA.....	104
3. La actividad transcripcional condiciona el perfil replicativo.....	106
Conclusiones	113
Bibliografía	117
Anexo I: Tablas suplementarias	135
Tabla de oligonucleótidos utilizados.....	137
Zonas excluidas del análisis.....	141
Anexo II: Listado de publicaciones	143

A word cloud background featuring various biological and scientific terms in different fonts and sizes, all in shades of gray. Prominent words include 'nucleosoma', 'replicación', 'transcripción', 'cromatina', 'iniciación', 'base', 'zonas', 'mayor', 'resultados', 'relación', 'genómico', 'manera', 'intermediario', 'análisis', 'durante', 'hecho', 'factores', 'iniciación', 'nucleosoma', 'mayor', 'resultados', 'relación', 'genómico', 'manera', 'intermediario', 'análisis', 'durante', 'hecho', 'factores', 'iniciación'.

RESUMEN

-

SUMMARY

La replicación del DNA consiste en la duplicación de la información genética codificada en el genoma de las células para poder ser transmitida a su descendencia. Este proceso debe producirse una única vez a lo largo del ciclo celular y se inicia en los orígenes de replicación. El reconocimiento de estos sitios de iniciación se realiza mediante complejos proteicos denominados replicadores cuyo mecanismo de acción está conservado desde levaduras, como *Saccharomyces cerevisiae*, donde el replicador reconoce una secuencia consenso, a células de metazoos en los que la especificidad de unión y actividad de los replicadores parece depender de una combinación de múltiples factores. El comienzo de la replicación en estos sistemas está estrechamente relacionado con otros procesos nucleares, especialmente con la transcripción génica, así como con la composición y estructura de la cromatina de las regiones en que se sitúan los orígenes de replicación.

Para profundizar en las relaciones entre estos procesos, hemos combinado una aproximación a escala genómica de los datos disponibles en la literatura correspondientes a marcas epigenéticas, motivos y estructuras secundarias de la secuencia de DNA en los orígenes de replicación más eficientes de sistemas celulares de mamíferos con estudios detallados a alta resolución de la estructura de la cromatina y la abundancia de intermediarios de replicación en orígenes específicos. De esta manera, hemos podido comprobar cómo estos orígenes eficientes se encuentran enriquecidos en variantes de histonas que forman nucleosomas lábiles así como en motivos de secuencia característicos de cromatina accesible que, en conjunto, facilitan el reclutamiento de los complejos de reconocimiento del origen así como del resto de factores que desencadenan la síntesis del DNA.

También hemos realizado análisis genómicos del perfil nucleosomal y de la abundancia de intermediarios de replicación de células promastigotes del parásito *Leishmania major*. La integración de estos resultados con datos del transcriptoma y de la localización de marcas epigenéticas disponibles en la literatura, nos ha permitido concluir que la transcripción activa es el determinante principal que subyace a la organización espacial y temporal de la replicación del genoma de este parásito. Estos resultados evidencian la naturaleza oportunista del proceso replicativo y sugieren que el acoplar el inicio de la replicación a la elongación de la transcripción podría ser una solución ancestral utilizada por las células eucariotas para replicar su material genético.

DNA replication consists in the duplication of the genetic information encoded within the cell genome for being transmitted to the next generation. This process must to occur once during the cell cycle and starts at the replication origins. Recognition of these initiation sites is performed by protein complexes called replicators, sharing similar mechanism from yeast -like *Saccharomyces cerevisiae* where the replicator recognizes one consensus sequence- to metazoan cells -where the binding specificity seems to depend on a combination of several factors. Replication initiation in these systems is highly related to other nuclear processes, specially transcription, and with the chromatin structure at those regions where replication origins are situated.

To go deeper into the relationship among these features, we have combined a genome-wide approach using published data for epigenetic marks, motifs and secondary structures of the DNA in the most efficient mammalian replication origins with high-resolution studies about the chromatin structure and the replication intermediates abundance at specific origins. In such a way, we have shown how these efficient origins are enriched in histone variants associated to labile nucleosomes and in sequence motifs characteristic for accessible chromatin. Both together supply the conditions for the recruitment of the origin recognition complexes and the rest of the protein factors that fire the DNA synthesis.

In addition, we have already performed genome-wide analysis of the nucleosomal profile and of the replication intermediates abundance using promastigotes from the human parasite *Leishmania major*. Integrating these results with the transcriptome and the location of some epigenetic marks leads to the conclusion that active transcription is the main determinant underlying the spatio-temporal organization of the replication in this system. These results evince the opportunistic nature under the replication of the DNA and suggest that the coupling of it to the transcription elongation may be an ancestral way to perform the replication of the entire genetic material.

Clave de abreviaturas utilizadas:

Ach3: Acetilación de histona **H3**

ACS: Secuencia consenso del ARS (del inglés: **ARS Consensus Sequence**)

ARS: Secuencia de replicación autónoma (del inglés: **Autonomous Replication Sequence**)

ATP: Trifosfato de adenosina (del inglés: **Adenosine Triphosphate**)

bp: Pares de bases (del inglés: **Base Pair**)

CDK: Quinasa dependiente de ciclina (del inglés: **Cyclin Dependent Kinase**)

CDS: Secuencia de DNA codificante (del inglés: **Coding DNA Sequence**)

CGI: Isla CpG (del inglés: **CpG Island**)

ChIP: Inmunoprecipitación de cromatina (del inglés: **Chromatin Immunoprecipitation**)

CMG: Complejo **Cdc45-MCM-GINS**

DDK: Quinasa dependiente de Dbf4 (del inglés: **Dbf4-dependent Kinase**)

DMEM: Medio de cultivo (del inglés: **Dulbecco's Modified Eagle's Medium**)

DMSO: Dimetil sulfóxido

DNA: Ácido desoxirribonucleico (del inglés: **Deoxyribonucleic Acid**)

DRIP: Inmunoprecipitación de DNA:RNA (del inglés: **DNA:RNA Immunoprecipitation**)

dsDNA: DNA de doble cadena (del inglés: **Double-Stranded DNA**)

EDTA: Ácido etilen diamino tetracético

EGTA: Ácido etilen glicol tetracético

ES: Células troncales embrionarias de ratón (del inglés: **Embryonic Stem**)

FBS: Suero fetal bovino (del inglés: **Fetal Bovine Serum**)

FT: Factor de Transcripción

G4: Cuádruplex de Guanina

GINS: Complejo Sld5-Psf1-3 (del japonés: **Go-Ichi-Ni-San**)

HEPES: Ácido 4-(2-hidroxyetil)-1-piperazin etano sulfónico

HT: Fin-inicio de transcripción en la misma hebra (del inglés: **Head-Tail**)

IR: Región intergénica (del inglés: **Intergenic Region**)

Kb: Kilobases

LIF: Factor inhibidor de leucemia (del inglés: **Leukemia Inhibition Factor**)

Mb: Megabases

MCM: Complejo de mantenimiento minicromosómico (del inglés: **Minichromosome Maintenance**)

MFA: Análisis de frecuencia de marcadores (del inglés: **Marker Frequency Analysis**)

MN: Mononucleosomas

NA: No Analizado

NDR: Región de baja ocupación nucleosomal (del inglés: **Nucleosome Depleted Region**)

NGS: Secuenciación de nueva generación (del inglés: Next-Generation Sequencing)	RNAPII: RNA Polimerasa II
nt: Nucleótido	RPKM: Lecturas por Kb por millón de alineamientos (del inglés: Reads per Kilobase per Million mapped reads)
NuSA: Ensayo de escaneo de nucleosomas (del inglés: Nucleosome Scanning Assay)	SDS: Dodecil sulfato sódico (del inglés: Sodium Dodecyl Sulfate)
ORC: Complejo de reconocimiento del ORI (del inglés: ORI Recognition Complex)	SL: Secuencia Líder
ORI: Origen de replicación	SNS: Cadenas nacientes pequeñas (del inglés: Short Nascent Strands)
PBS: Tampón fosfato salino (del inglés: Phosphate Buffer Saline)	SSC: Citrato sódico salino (del inglés: Saline Sodium Citrate)
PCR: Reacción en cadena de la polimerasa (del inglés: Polymerase Chain Reaction)	ssDNA: DNA de cadena sencilla (del inglés: Single-Stranded DNA)
PNK: Polinucleótido Kinasa	SSR: Región de cambio de cadena (del inglés: Strand Switch Region)
pre-IC: Complejo de pre-iniciación (del inglés: pre-Initiation Complex)	SSRc: SSR Convergente
pre-RC: Complejo pre-replicativo (del inglés: pre-Replicative Complex)	SSRd: SSR Divergente
PTU: Unidad de transcripción policistrónica (del inglés: Polycistronic Transcription Unit)	TE: TRIS-EDTA
qPCR: PCR cuantitativa (del inglés: quantitative PCR)	TEN: Tampón TRIS-EDTA-NaCl
RNA: Ácido ribonucleico (del inglés: Ribonucleic Acid)	TRIS: Tris(hidroximetil) aminometano
RNAi: RNA interferente	TSS: Sitio de comienzo de la transcripción (del inglés: Transcription Start Site)
RNAm: RNA mensajero	TTS: Sitio de terminación de la transcripción (del inglés: Transcription Termination Site)
	U: Unidades enzimáticas
	UTR: Regiones no traducidas (del inglés: Untranslated Region)



INTRODUCCIÓN

1. Replicación del DNA

1.1. Interés biológico y progreso histórico

La cromatina es el conjunto de DNA y proteínas que almacena la información genética contenida en el núcleo de cada célula eucariota. Como paso previo a la división celular, esta información debe copiarse de manera íntegra y única en cada ciclo celular para su idéntico reparto entre las dos nuevas células hijas. La duplicación de este material genético se conoce como replicación del DNA y comienza en regiones llamadas orígenes de replicación (ORI).

Dada su vital importancia y el elevado número de veces que se tiene que producir -tanto a nivel poblacional en organismos unicelulares como a nivel celular en pluricelulares-, la iniciación de la replicación constituye uno de los procesos celulares más y mejor regulados. Las alteraciones del mismo conducen a graves trastornos y enfermedades. A lo largo de varios trabajos, el grupo del Prof. Andrew Jackson ha descrito que, mutaciones en algunos de los componentes de los complejos proteicos que se unen a los ORI provocan graves problemas en el crecimiento y en el desarrollo de modelos como el pez cebra y, en humanos, son responsables de los defectos del síndrome de Meier-Gorlin (Bicknell *et al.*, 2011; de Munnik *et al.*, 2012). Además, las disfunciones en la iniciación de la replicación en los ORI o en los procesos de control de su correcta activación durante la fase S del ciclo celular conllevan estrés replicativo que puede contribuir al desarrollo tumoral así como a procesos de inmunodeficiencias y ciertas enfermedades raras (revisado en Muñoz y Méndez, 2016). De hecho, como estrategia para hacer frente a muchas de estas patologías se ha postulado la inhibición de las polimerasas implicadas en esa replicación del DNA desregulada (revisado en Berdis, 2008) y, del mismo modo, las diferencias de los componentes del replisoma de bacterias respecto a su hospedador pueden ser utilizadas como diana terapéutica antibacteriana (revisado en Robinson *et al.*, 2012). Por otra parte, el mantenimiento de la conformación de la cromatina tras el avance de las horquillas de replicación es igualmente crítico: alteraciones en la información epigenética pueden derivar en cáncer (revisado en Sharma *et al.*, 2010), inestabilidad en las horquillas de replicación (Clemente-Ruiz y Prado, 2009) u otras patologías asociadas al estrés replicativo (revisado en Alabert y Groth, 2012). Por todo ello, resulta evidente el interés por el estudio y la ampliación del conocimiento de los procesos y mecanismos que especifican el inicio de la replicación en genomas complejos así como de la organización de la cromatina en los mismos.

Hace poco más de medio siglo, se expuso la teoría del replicón, que sigue siendo la base sobre la que se sustenta el conocimiento actual acerca del comienzo de la replicación del DNA. Este paradigma define como replicador a la secuencia o región del DNA que es reconocida por un elemento *trans* o iniciador, generando la apertura de la doble hélice y por tanto, el comienzo de todo el proceso (Jacob *et al.*, 1963). A finales de esa década y mediante el análisis de fibras de DNA humano marcadas radiactivamente, se observaron por primera vez estos replicones; regiones en las que se sintetizaba el DNA de manera bidireccional e independientemente unas de otras. Ese mismo trabajo también sirvió para establecer la tasa de replicación y calcular el número de ORI por célula (Huberman y Riggs, 1968). Durante los años posteriores, se presentaron diversos estudios de caracterización y aislamiento de replicones en modelos como *Drosophila melanogaster* (Blumenthal *et al.*, 1974), *Saccharomyces cerevisiae* (revisado en Stinchcomb *et al.*, 1979), *Xenopus* (Mechali y Kearsley, 1984) y células de mamíferos (Heintz y Hamlin, 1986; Nakamura *et al.*, 1986). Poco después, utilizando *S. cerevisiae* como modelo de estudio, el laboratorio del Prof. Bruce Stillman comenzó a describir las primeras proteínas que actuaban como iniciadores (Diffley y Stillman, 1988) hasta la identificación del complejo de reconocimiento del origen (ORC) como el factor que reconocía al replicador (Bell y Stillman, 1992). A partir de los años 90, destacaron los trabajos, cada vez más precisos, sobre ORI específicos así como la caracterización de los elementos implicados en el reconocimiento, activación y regulación de los mismos, desde modelos de levaduras (Dubey *et al.*, 1996; Gómez y Antequera, 1999) a organismos pluricelulares como *Drosophila* (Austin *et al.*, 1999) o vertebrados (Abdurashidova *et al.*, 2000; Prioleau *et al.*, 2003). Finalmente, en esta década, con el auge de las técnicas de hibridación y fluorescencia primero (*microarray*) y de secuenciación masiva (NGS) después, el campo de estudio de la replicación del DNA se ha dirigido en gran parte a la caracterización a escala genómica de los eventos de inicio de replicación en distintos modelos (Wyrick *et al.*, 2001; Cadoret *et al.*, 2008; Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011; Costas *et al.*, 2011; Besnard *et al.*, 2012; Comoglio *et al.*, 2015).

Una de las conclusiones que se extraen de todos estos estudios es que el número de ORI resulta proporcional a la longitud del genoma eucariótico (**Figura 1**). Desde el caso de bacterias y arqueas con cromosomas muchas veces únicos, circulares y pequeños donde suele existir un solo ORI -en algunos casos, unos pocos- (revisado en Barry y Bell, 2006) a modelos como ratón con aproximadamente 50.000 ORI deducidos a partir de las distancias inter-ORI observadas en estudios de hibridación a *microarray* de una fracción significativa de los 2.700 Mb del genoma

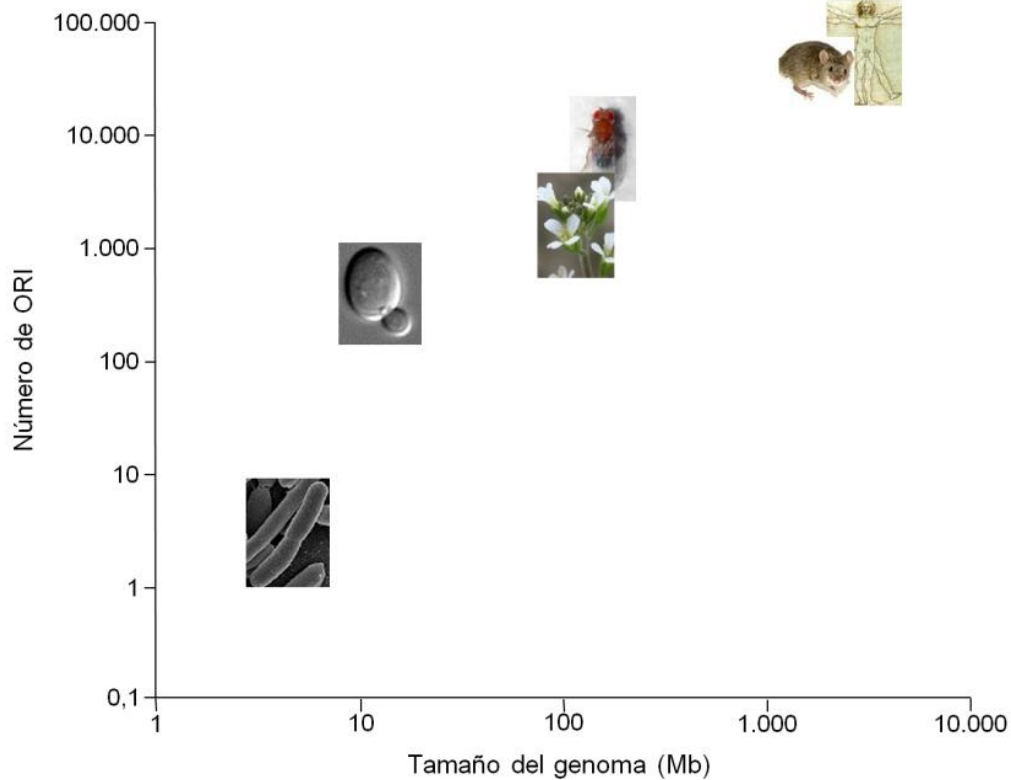


Figura 1: Relación entre el tamaño del genoma (en Mb) y el número de ORI estimado por célula en cada uno de los sistemas estudiados: bacterias (revisado en Barry y Bell, 2006), levaduras (Nieduszynski *et al.*, 2006; Feng *et al.*, 2006), *Arabidopsis thaliana* (Costas *et al.*, 2011), *D. melanogaster* (Comoglio *et al.*, 2015), ratón (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2015) y humano (Huberman y Riggs, 1968; Picard *et al.*, 2014).

(Sequeira-Mendes *et al.*, 2009) o mediante secuenciación de pequeños intermediarios de replicación (SNS) -SNS-seq- (Cayrou *et al.*, 2015). No obstante, proporcionalmente al aumento de la complejidad del genoma, también crece la dificultad para delimitar el número de ORI correctamente. Un caso paradigmático que refleja este problema es el sistema de células humanas en el que, con el aumento de la sensibilidad de las técnicas a escala genómica en el último lustro, se han llegado a postular hasta 300.000 ORI (Besnard *et al.*, 2012) aunque este dato se considera sobreestimado y el número podría situarse más correctamente alrededor de 50.000-100.000 (Picard *et al.*, 2014), un rango de valores más coherente con los 30.000-50.000 descritos inicialmente mediante marcajes radiactivos de fibras de DNA (Huberman y Riggs, 1968). Sin embargo, considerar activados todos los ORI caracterizados a nivel poblacional -mediante NGS-, arroja valores cinco veces superiores a los definidos a partir del análisis de moléculas individuales (Cayrou *et al.*, 2011; Kaykov *et al.*, 2016). Esta variabilidad en las estimaciones de la densidad de ORI revela la elevada

flexibilidad del programa replicativo: existe un gran número de ORI potenciales en el genoma y cada célula activa una cohorte diferente de ellos; es decir, un modelo en el que “muchos son llamados (licenciados) pero pocos son elegidos (activados)” (propuesto en DePamphilis, 1996 y revisado en DePamphilis, 1999). A lo largo de esta Introducción, profundizaremos en esta idea.

1.2. Metabolismo de la iniciación de la replicación del DNA en eucariotas: licenciamiento y activación

El proceso bioquímico de ensamblaje de los componentes que desencadenan la iniciación de la síntesis del DNA en los ORI está altamente conservado en los distintos sistemas eucariotas analizados, tanto funcional como estructuralmente, así como también su sincronización con el ciclo celular. Este conocimiento, detallado a continuación, ha sido ampliamente revisado entre otros en Méndez y Stillman (2003); DePamphilis (2005); Aladjem (2007); Masai *et al.* (2010); Diffley (2011); Musiacke y Rybaczek (2015); y Fragkos *et al.* (2015).

Antes de la mitosis, en la transición entre las fases G2 a M (G2/M) del ciclo celular, el complejo ORC, formado por seis subunidades (nombradas Orc1 a Orc6), se une a aquellas regiones que tienen potencial para disparar el inicio de la replicación; los ORI. Durante la fase G1, el complejo ORC recluta los factores de iniciación Cdc6 y Cdt1. Este segundo factor es necesario para la unión de la proteína Cdc6 que, junto a la subunidad Orc1, presentan la actividad ATPasa responsable de la unión de otros dos grupos proteicos, los complejos de mantenimiento minicromosómicos o MCM. Esta unión también está favorecida por la actividad del factor Cdk2 (Powell *et al.*, 2015). Cada uno de los complejos MCM está formado por 6 subunidades (denominadas Mcm2 a Mcm7) que se reclutan asimétricamente respecto al complejo ORC (Evrin *et al.*, 2009; Remus *et al.*, 2009; Gambus *et al.*, 2011 y revisado en Lutzmann y Méchali, 2009). Los complejos MCM permanecen asociados a las horquillas de replicación una vez que el ORI se activa, ya que su actividad helicasa es imprescindible para el avance de éstas (Labib *et al.*, 2000). La unión de los complejos Mcm2-7 al DNA completa el ensamblaje de los llamados complejos pre-replicativos (pre-RC) o licenciamiento del ORI durante la fase G1 del ciclo celular (**Figura 2**).

Durante la transición entre las fases G1 a S (G1/S) se produce la activación del subgrupo de ORI que van a iniciar la síntesis del DNA. Esta cascada de activación co-

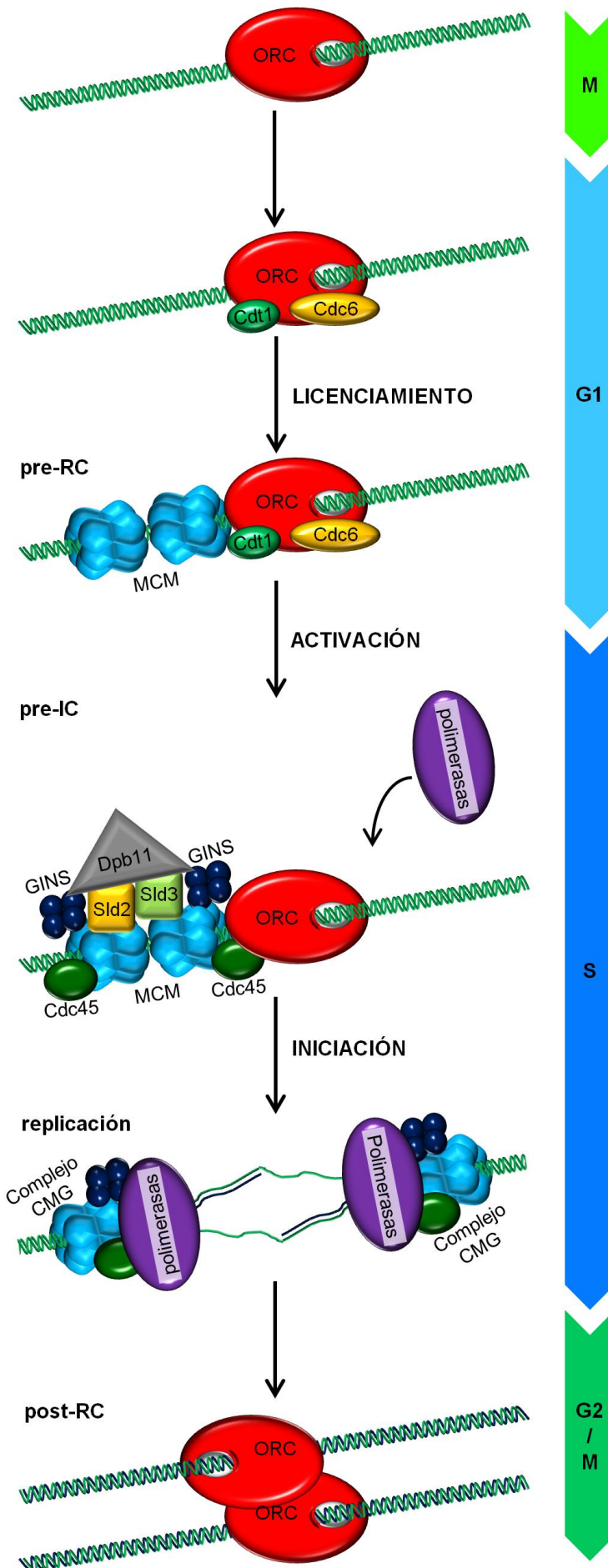


Figura 2: Esquema simplificado del licenciamento y activación de un ORI eucariótico.

ORC reconoce un *locus* capaz de iniciar la replicación. Durante la fase M/G1, se unen los factores de iniciación Cdt1 y Cdc6 que reclutan dos complejos MCM formando el pre-RC. El licenciamento del origen, conlleva la posibilidad (aunque no la necesidad) de que éste se active.

Aquellos ORI que se activan experimentan una cascada de fosforilaciones sucesivas al inicio de la fase S en las que los componentes del pre-RC -Cdt1 y Cdc6- son sustituidos por otros factores como Cdc45, Sld2, Sld3, que también serán fosforilados permitiendo el reclutamiento del tetrámero GINS completando el pre-IC. A su vez, esto permite la introducción de las polimerasas, la activación de las helicetas (complejo CMG) y el inicio de la replicación.

Finalmente, tras la replicación bidireccional, los ORI vuelven a ser reconocidos por ORC, impidiendo su activación hasta el siguiente ciclo celular

mienza cuando las quinasas dependientes de Dbf4 (DDK) fosforilan las subunidades Mcm4 y Mcm6 del complejo MCM (Randell *et al.*, 2010; Sheu y Stillman, 2010) lo que conlleva la sustitución de la proteína Cdc6 por Cdc45 y el reclutamiento de los factores Sld2 y Sld3. Posteriormente, las quinasas dependientes de ciclinas (CDK) fosforilan estos factores (Zegerman y Diffley, 2007) permitiendo el reclutamiento de componentes adicionales de la activación como Mcm10, Dpb11 y el tetrámero formado por las subunidades Sld5 y Psf1-3 (complejo GINS) que constituyen el complejo de preiniciación o pre-IC (Gambus *et al.*, 2006; Pacek *et al.*, 2006; Gambus *et al.*, 2009) (**Figura 2**). La unión de estos factores promueve la actividad helicasa del complejo CMG (Cdc45-MCM-GINS) y, en último término, tras el reclutamiento de las diferentes polimerasas así como del resto de componentes del replisoma, desencadena el inicio de la síntesis del DNA (Ilves *et al.*, 2010; Costa *et al.*, 2011; y revisado en Boos *et al.* 2012) (**Figura 2**).

En los últimos años se han producido enormes avances en la reconstitución bioquímica del proceso de licenciamiento de los ORI eucariotas. Por ejemplo, se ha conseguido reconstituir *in vitro* los complejos pre-RC aislados de extractos de *S. cerevisiae* en las fases G1 y S del ciclo celular (Gros *et al.*, 2014). Más recientemente, se ha reproducido el proceso completo de activación de los complejos pre-IC en DNA a partir de la purificación de los factores proteicos ORC, Cdc6, Cdt1-Mcm2-7, DDK, CDK, Sld2, Sld3, Cdc45, Dpb11, Mcm10, GINS y la polimerasa ϵ (Pol ϵ), demostrando que estos son los componentes suficientes y necesarios para la iniciación de la replicación del DNA (Yeeles *et al.*, 2015).

La correcta regulación de la iniciación de la síntesis de DNA es fundamental para que la copia de la información genética sea única en el ciclo celular, evitando inicios adicionales de la replicación del DNA (re-replicación). En este punto destaca el papel regulador de las CDK -cuya concentración aumenta en la transición de fases G1/S- no sólo disparando la activación de los ORI licenciados mediante la fosforilación de Sld2 y Sld3 (Zegerman y Diffley, 2007), sino también inhibiendo la formación de nuevos complejos pre-RC (Dahmann *et al.*, 1995) y confinando el licenciamiento a la fase M/G1 -donde la concentración de CDK es menor- (revisado en Diffley, 2004).

1.3. Orígenes de replicación en diversos modelos: especiación

En bacterias, la secuencia del replicador, llamada *oriC*, presenta una elevada composición en A/T y suele contener secuencias conservadas que son reconocidas

por el iniciador (DnaA) que separa las hebras de DNA en la región del ORI permitiendo la unión de dos helicasas (DnaB) alrededor de cada una de ellas (revisado en Zakrzewska-Czerwinska *et al.*, 2007; Leonard y Grimwade, 2015). Además de este mecanismo canónico existente en el cromosoma bacteriano, en el plásmido ColE1 de *Escherichia coli*, la iniciación de la replicación ocurre a través de la transcripción de un RNA que forma un híbrido DNA:RNA actuando como cebador de la síntesis de la nueva cadena de DNA (Masukata y Tomizawa, 1990). En la mitocondria, orgánulo con evidentes paralelismos con procariontes, también se han descrito acoplamientos entre la iniciación de la replicación y la transcripción por medio de estructuras híbridas DNA:RNA (*R-loops*) (Chang *et al.*, 1985).

La iniciación de la replicación en arqueas muestra un comportamiento intermedio entre bacterias y eucariotas. Si bien existen uno o varios *oriC* por cada cromosoma -especialmente cuando son circulares y pequeños- y el iniciador comparte rasgos con el DnaA bacteriano, este también exhibe dominios similares a las proteínas Orc1 y Cdc6 de eucariotas (revisado en Barry y Bell, 2006; Wigley, 2009; Kelman y Kelman, 2014). Curiosamente, en eucariotas como la familia de los tripanosomátidos (donde destacan los géneros *Trypanosoma* y *Leishmania*), el iniciador también es una proteína única que presenta motivos homólogos tanto a Orc1 como a Cdc6 (Godoy *et al.*, 2009).

S. cerevisiae es uno de los principales modelos de estudio del inicio de la replicación en eucariotas. En esta levadura, los ORI se identificaron como secuencias capaces de conferir replicación autónoma en plásmidos (ARS) (Marahrens y Stillman, 1992; revisado en Stinchcomb *et al.*, 1979). En los cromosomas, las ARS se extienden a lo largo de 100-150 bp y presentan un enriquecimiento en A/T que contiene la secuencia consenso (ACS) de 11 bp: (A/T)TTTA(T/C)(A/G)TTT(A/T) que es imprescindible, aunque no suficiente, para la actividad de iniciación (revisado en Newlon y Theis, 1993). Esta ACS es reconocida específicamente por el complejo ORC (Bell y Stillman, 1992). Las ARS también pueden contener elementos adicionales más o menos conservados y ricos en A/T (B1, B2 y B3) y que tienen como función la unión de diversos factores que incrementan la actividad del ORI, incluyendo en algunos casos, factores de transcripción (Rao *et al.*, 1994). En la levadura de fisión *Schizosaccharomyces pombe*, las regiones ARS se extienden a lo largo de una mayor longitud, entre 500 y 1000 bp y carecen de una secuencia consenso equivalente a la ACS de *S. cerevisiae*. No obstante, estos replicadores presentan dominios enriquecidos en A/T (islas A/T) que son específicamente reconocidos por el dominio A/T-hook de la subunidad Orc4 de este organismo (Chuang y Kelly, 1999; revisado en

Mojardin *et al.*, 2013). Curiosamente, aunque las islas A/T de *S. pombe* se sitúan en regiones intergénicas, tienden a localizarse en las cercanías de promotores, sugiriendo una relación con la transcripción (Gómez y Antequera, 1999; Segurado *et al.*, 2003).

La identificación de los ORI a través de los ensayos de mantenimiento de plásmidos no ha podido ser extrapolada a eucariotas pluricelulares. De hecho, en modelos como *Xenopus* (Méchali y Kearsey, 1984) o células humanas (Heinzel *et al.*, 1991), la capacidad de replicar una molécula de DNA está más relacionada con la longitud de la misma que con la secuencia. Aunque los ORI identificados en *Xenopus laevis* presentan cierta preferencia por regiones ricas en A/T (Stanojic *et al.*, 2008), los ORI de células de mamíferos caracterizados en detalle, como los situados en la proximidad de los promotores de los genes *β -globina* (Kitsberg *et al.*, 1993), *HSP70* (Taira *et al.*, 1994) o *PPV1* (Giacca *et al.*, 1994), no sólo no están enriquecidos en A/T sino que se asocian con regiones ricas en C/G. Estudios más recientes de localización genómica de ORI en células de ratón y de *Drosophila* mediante el aislamiento y mapeo de SNS han permitido asociar más del 90% y del 65% de los ORI respectivamente, a elementos ricos en guanina (Cayrou *et al.*, 2012). Gran parte de estos elementos tienen el potencial de formar cuádruplex de guanina (G4), estructuras secundarias en la cadena de DNA estabilizadas por medio de puentes de hidrógeno entre cuatro guaninas situadas en el mismo plano. La asociación de los ORI con secuencias potencialmente formadoras de G4 también se ha establecido en células humanas (Besnard *et al.*, 2012). Además, experimentos de mutagénesis dirigida contra los elementos necesarios para la formación del G4 en dos ORI particulares estudiados en células de *Gallus gallus* muestran que se produce un descenso de la eficiencia de iniciación de la replicación (Valton *et al.*, 2014). Todos estos resultados sugieren que la riqueza en G/C parece ser lo más parecido a una secuencia consenso del replicador en eucariotas pluricelulares. Es interesante que en el modelo vegetal *A. thaliana*, cuyo genoma exhibe una alta composición en A/T, el estudio de los ORI a escala genómica muestra incluso una preferencia de los mismos por regiones ricas en C/G (Costas *et al.*, 2011). Además de elementos intrínsecos a la estructura del replicón como motivos de secuencia o estructuras secundarias, el inicio de la replicación en sistemas como animales y plantas también se ve influenciado por la transcripción o la estructura de la cromatina alrededor de los ORI, como se desarrollará en el segundo epígrafe de la Introducción.

1.4. Análisis de la iniciación de la replicación en poblaciones celulares: eficiencia de activación de los orígenes de replicación

En los estudios del inicio de replicación con cultivos celulares, hay que diferenciar dos conceptos. Por una parte, los ORI que se activan a nivel de cada célula individual y, por otra, la contribución de cada uno de ellos al perfil replicativo de la población analizada.

A nivel celular, los ORI licenciados son aquellos en los que la formación del complejo pre-RC habilita que puedan activarse en la fase S. Sin embargo, sólo una pequeña fracción de ellos lo hace. La utilización de unos u otros viene determinada por diversos factores como el programa temporal -regiones de replicación temprana o tardía-, la estructura de la cromatina en la que se encuentra el ORI -eucromatina o heterocromatina- o el tipo celular (revisado en DePamphilis, 1999; Bielinsky, 2003; Aladjem, 2007). El exceso en el número de ORI potenciales resulta una ventaja para hacer frente a condiciones de estrés replicativo que pudieran conducir hacia la infrarreplicación del genoma. De hecho, la reducción de los complejos MCM mediante técnicas de RNAi en *Caenorhabditis elegans* y en células humanas no resulta letal bajo condiciones normales, al contrario de lo que ocurre cuando además se induce en ellas un estrés replicativo (Woodward *et al.*, 2006; Ibarra *et al.*, 2008). Del mismo modo, en ensayos en extractos de huevos de *X. laevis*, se ha comprobado que el exceso de complejos MCM puede activar diversos ORI “durmientes” que en condiciones normales se mantendrían inactivos (Woodward *et al.*, 2006).

Así pues, en función de su activación se pueden definir tres tipos de ORI: constitutivos, que se activan en todas las células de la población; inactivos, que aunque licenciados, no se activan en ningún individuo y flexibles, aquellos ORI que se activan de manera estocástica y alternativamente en unas células pero no en otras y que contribuyen en conjunto al perfil replicativo de un dominio determinado del genoma (**Figura 3** y revisado en Méchali, 2010). Un ejemplo paradigmático de estos grupos de ORI se encuentra en el *locus DHFR* de hámster donde se han descrito hasta veinte sitios de inicio de la replicación diferentes (Dijkwel *et al.*, 2002). En el plano poblacional, la activación desigual de los ORI entre las células provoca diferencias en la detección de aquellos que se activan en mayor número de individuos respecto a los que lo hacen en pocos casos (**Figura 3**). Por tanto, se considera a los primeros como ORI eficientes, definiendo esta eficiencia como la frecuencia de activación de un ORI en la población celular.

La eficiencia de un ORI es la resultante de diversos factores que, a su vez, pueden estar interrelacionados entre ellos. Por ejemplo, se ha observado que los ORI más eficientes en células de mamífero se sitúan en promotores (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011; Besnard *et al.*, 2012) o están asociados a determinadas secuencias como las que presentan capacidad de formar estructuras G4 (Cayrou *et al.*, 2012; Valton *et al.*, 2014). Es también en estas regiones, debido a una organización menos condensada de la cromatina, donde se produce el acceso de diversos factores proteicos al DNA. Por tanto, como se desarrolla en más profundidad en el siguiente apartado, la eficiencia de activación de los ORI también está íntimamente relacionada con la transcripción y la conformación de la cromatina.

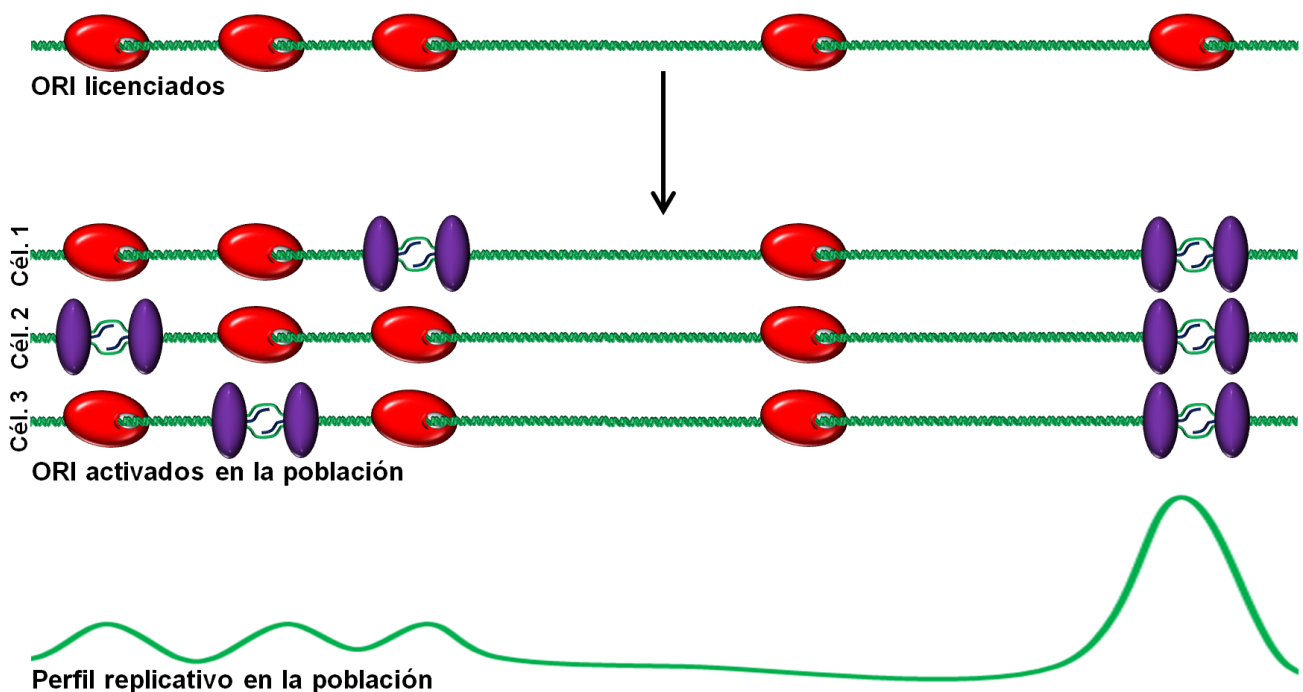


Figura 3: Esquema de la activación de los ORI en una población celular. En la parte superior, representación de los ORI licenciados en toda la población, es decir, todos aquellos con posibilidad de activación. A continuación, representación de los tres tipos de ORI que se pueden encontrar en base a su activación: (de derecha a izquierda) constitutivos -se activan en toda la población-, inactivos o “durmientes” -en condiciones normales no se activan en ninguna de las células- o flexibles -se activan en unas células pero no en otras; además, suelen aparecer agrupados-. En la parte inferior, representación del perfil replicativo reflejando la contribución de cada origen individual al perfil que se obtendría del estudio completo de la población.

2. Cromatina y regulación genómica

El nucleosoma es la unidad básica de la cromatina. Los nucleosomas canónicos están compuestos por dos unidades de cada una de las histonas H2A, H2B, H3 y H4 conformando un octámero alrededor del cual se enrollan de promedio 147 bp de DNA dando 1,7 vueltas (revisado en Kornberg, 1974). Sin embargo, existen variantes de estas histonas que pueden reemplazar a alguna de las subunidades en ciertas localizaciones genómicas y que, junto con las modificaciones químicas de los aminoácidos de las colas de las histonas, juegan un papel clave, como una especie de “código”, en la regulación de la transcripción, replicación y otros procesos bioquímicos del material genético (revisado en Jenuwein y Allis, 2001). En la **Figura 4** se muestra un diagrama que resume las características compartidas por estos procesos nucleares.

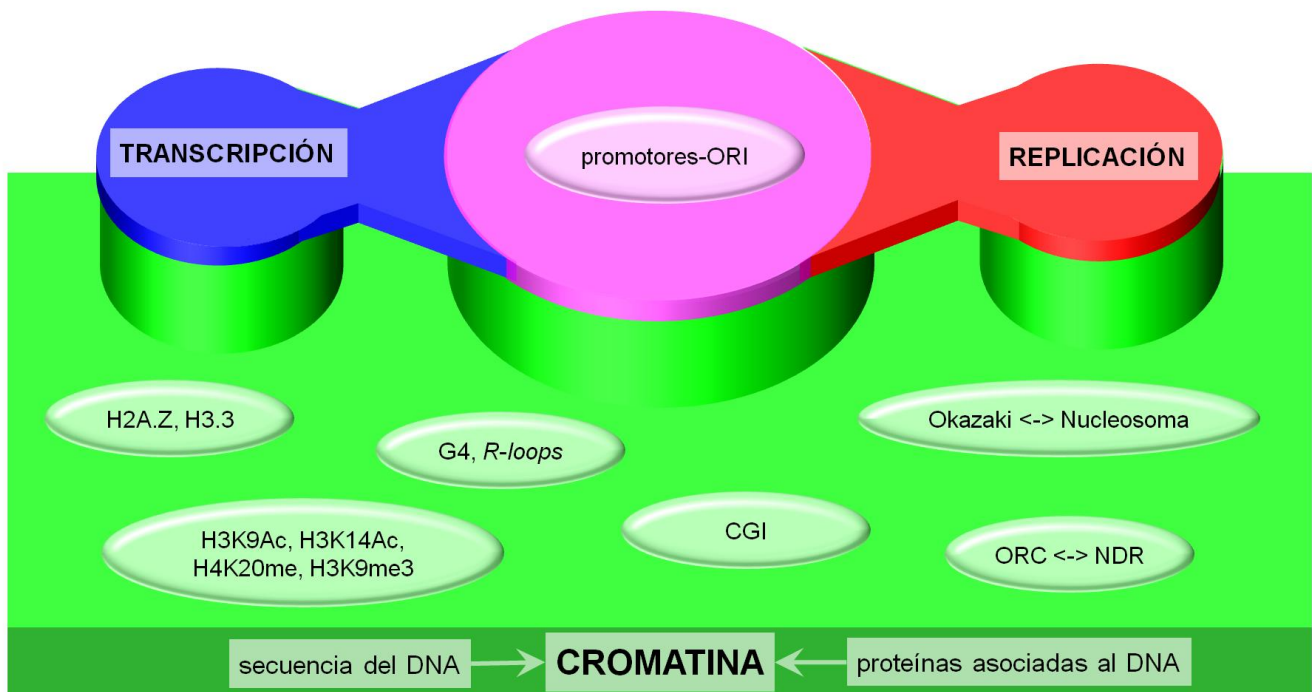


Figura 4: Diagrama que representa cómo la interrelación entre los procesos nucleares de replicación y transcripción se sustentan sobre distintas características determinadas por la cromatina.

2.1. Estructura de la cromatina y expresión génica

Un alto grado de empaquetamiento de la cromatina limita, por impedimento estérico, la unión de factores proteicos al mismo, incluyendo el acceso de la RNAPII a los promotores de los genes y, por tanto, inhibiendo el proceso de transcripción (revisado en Kornberg y Lorch, 1999). De hecho, las zonas más expuestas -menor condensación de la cromatina- coinciden con regiones reguladoras de la transcripción en diversos modelos de estudio como *Drosophila* (Farkas *et al.*, 2000), levaduras (Lee *et al.*, 2007) o células de mamíferos (Schones *et al.*, 2008). En estudios recientes realizados a muy alta resolución, se han localizado a escala genómica partículas subnucleosomales correspondientes probablemente a factores de transcripción unidos al DNA en regiones no cubiertas por nucleosomas (Henikoff *et al.*, 2011). Es más, se ha demostrado una relación muy estrecha entre el posicionamiento de los factores de transcripción y la reposición de los propios nucleosomas (Li *et al.*, 2015 y revisado en Maheara y Ohkawa, 2015).

La incorporación en los nucleosomas de algunas variantes de las histonas canónicas como H2A.Z o H3.3 (variantes de las histonas H2A y de H3, respectivamente), confieren una mayor inestabilidad a la unión del nucleosoma con el DNA (Jin y Felsenfeld, 2007). De esta forma, su presencia favorece la tasa de recambio del octámero, lo que facilita la competición con los factores de transcripción u otros complejos proteicos que tienen mayor probabilidad de unirse al DNA en estas regiones. De hecho, las regiones de los promotores o de los *enhancers* que se han considerado clásicamente como libres de nucleosomas están en realidad ocupadas por nucleosomas lábiles (formados por estas variantes de histonas), que las metodologías de estudio empleadas habitualmente no consiguen preservar (Jin *et al.*, 2009).

Los extremos N-terminal de las histonas también pueden ser modificados por la adición de grupos químicos como acetilo, metilo, ubiquitina o fosfato que alteran la estabilidad del nucleosoma y/o facilitan el reclutamiento de diversos factores. Por ejemplo, la acetilación de las lisinas 9 y/o 14 de la histona H3 (H3K9Ac y H3K14Ac) aparece en regiones de promotores activos (Koch *et al.*, 2007). También lo hace la metilación de la lisina 20 de H4 (H4K20me) o la trimetilación de la lisina 4 (H3K4me3) de H3. Sin embargo, metilaciones en las lisinas 9 y/o 27 de H3 (H3K9me3 y H3K27me) están asociadas a represión de la transcripción (Barski *et al.*, 2007). También las histonas presentes en secuencias de *enhancers* pueden sufrir

modificaciones, como la acetilación de la lisina 27 de H3 (H3K27Ac) que marca la diferencia entre los activos y los inactivos (Creighton, 2010).

Otro tipo de modificación química relacionada con la represión de la transcripción es la metilación de las citosinas cuando estas preceden a guanina (dinucleótidos CpG). Se calcula que más de dos tercios de las secuencias CpG del genoma están metiladas en mamíferos, impidiendo que la transcripción comience en esas regiones (Ehrlich *et al.*, 1982). Sin embargo, los dinucleótidos CpG no metilados presentan una importancia crucial. Normalmente, estos se agrupan en regiones llamadas islas CpG (CGI), que se extienden a lo largo de una longitud aproximada de 1 Kb, donde el contenido en C/G es superior respecto a la media genómica y presentan una estructura de la cromatina diferenciada (Bird *et al.*, 1985; revisado en Tazi y Bird, 1990). Las CGI son el lugar de unión de muchos factores de transcripción y de hecho, más del 60% de los sitios de inicio de la transcripción (TSS) de células humanas se encuentran asociados a las mismas (Cuadrado *et al.*, 2001 y revisado en Antequera, 2003). Como corresponde a regiones promotoras, es habitual la presencia de modificaciones de histonas relacionadas con la activación de la transcripción como H3K9Ac y H3K14Ac (Koch *et al.*, 2007). También existen evidencias de una asociación de la metilación en CpG con la de algunas modificaciones de histonas sugiriendo mecanismos interrelacionados entre ambos procesos (Fouse *et al.*, 2008; Illingworth *et al.*, 2010).

2.2. Estructura de la cromatina y replicación del DNA

La mayor accesibilidad de la cromatina en ciertas regiones genómicas para favorecer el comienzo de la transcripción sugiere que esta conformación podría ser aprovechada por otros procesos nucleares que requieren el acceso al DNA; uno de ellos sería la propia replicación. Esta posibilidad se empezó a considerar por primera vez al comprobar que las regiones de replicación temprana coincidían con muchos promotores de genes constitutivos en ratón (Goldman *et al.*, 1984). En células de *Drosophila* se ha descrito una alta correspondencia entre los sitios de unión de ORC y los promotores de los genes en dominios de replicación temprana (MacAlpine *et al.*, 2004). Resultados similares se han obtenido en otros sistemas. Como se ha comentado anteriormente, en *S. pombe*, por ejemplo, los ORI tienden a localizarse en regiones intergénicas cercanas a promotores (Gómez y Antequera, 1999). Asimismo, los primeros ORI que se caracterizaron en células de mamíferos -en los promotores de los genes de β -globina (Kitsberg *et al.*, 1993), *HSP70* (Taira *et al.*, 1994) y *PPV1*

(Giacca *et al.*, 1994; Abdurashidova *et al.*, 2000)- coinciden con los TSS. Las CGI, a las que se mantienen asociados buena parte de los promotores en células de mamífero, son otro excelente ejemplo de acoplamiento entre transcripción y replicación. Al aislar intermediarios de replicación enriquecidos en la proximidad del ORI en células humanas, se demostró que presentaban un alto grado de secuencias provenientes de CGI (Delgado *et al.*, 1998). Por otra parte, experimentos de inmunoprecipitación de cromatina (ChIP) con anticuerpos específicos contra las subunidades ORC1 y ORC2 en células humanas identificaron numerosos ORI en CGI (Keller *et al.*, 2002; Ladenburger *et al.*, 2002; Dellino *et al.*, 2013). Además, los ORI constitutivos más conservados entre los diferentes tipos celulares de ratón y humano están asociados a promotores de CGI (Sequeira-Mendes *et al.*, 2009; Besnard *et al.*, 2013; Cayrou *et al.*, 2015).

Teniendo en cuenta la relación descrita entre los TSS y los ORI, así como entre los primeros y la estructura de la cromatina, parece lógico pensar que el paisaje nucleosomal también tenga un importante papel en la definición y eficiencia del inicio de la replicación. En *S. cerevisiae*, se ha demostrado que la unión de ORC a la secuencia ARS1 es necesaria para el correcto posicionamiento de los nucleosomas que lo rodean (Lipford y Bell, 2001). Es más, estudios a escala genómica corroboran que ORC se localiza en la región libre de nucleosomas del ARS (Eaton *et al.*, 2012). También en células de *Drosophila* se ha comprobado que ORC se une a regiones libres de nucleosomas (MacAlpine *et al.*, 2010).

Curiosamente, en otros modelos, como *S. pombe*, los sitios de unión del complejo ORC no coinciden con regiones libres de nucleosomas (de Castro *et al.*, 2012) y, en *A. thaliana*, las regiones a las que se une ORC están enriquecidas en secuencias en las que se predice una alta ocupación nucleosomal (Stroud *et al.*, 2012). La misma predicción se ha encontrado analizando las secuencias de los ORI de ratón (Cayrou *et al.*, 2012). Esta aparente contradicción, al igual que en el caso de la relación entre cromatina y transcripción comentada anteriormente (Jin *et al.*, 2009), podría ser explicada por razones metodológicas puesto que los nucleosomas inestables -que biológicamente podrían desarrollar un papel en el reconocimiento y unión de ORC al DNA- difícilmente son capturados por los protocolos habituales de preparación de cromatina.

Dejando a un lado la controversia sobre si el complejo ORC se une a regiones del DNA empobrecidas en nucleosomas, en *S. cerevisiae* se ha descrito otra interesante interacción entre el proceso de replicación y la conformación de la

cromatina. Combinando elegantes estudios genéticos y genómicos se ha demostrado que el reposicionamiento de los nucleosomas en la hebra de DNA recién replicada está acoplado a la síntesis de los fragmentos de Okazaki (Smith y Whitehouse, 2012).

Recientemente, la integración de los datos disponibles de modificaciones y variantes de histonas, localización de ORI o de complejos como ORC o Polycomb, sobre los que se han aplicado diversos modelos estadísticos, ha supuesto un paso adelante en el estudio de la relación entre varios procesos nucleares y los estados de la cromatina. En *Drosophila* se describieron originalmente cinco estados de cromatina (Filion *et al.*, 2010), que posteriormente se aumentaron hasta nueve incorporando al análisis una mayor cantidad de datos genómicos (Kharchenko *et al.*, 2011). También en *A. thaliana* se han realizado diversos esfuerzos dirigidos a caracterizar motivos diferenciados en la cromatina. Hace unos años, con la información disponible en aquel momento, se identificaron cuatro dominios principales correspondientes a heterocromatina, unión de Polycomb, genes activos y regiones intergénicas (Roudier *et al.*, 2011). Estos estados han sido recientemente ampliados a nueve diferentes (Sequeira-Mendes *et al.*, 2014). Siguiendo esta metodología e integrando hasta 43 marcas genómicas diferentes junto con una caracterización de regiones de inicio de la replicación a partir de SNS, se ha postulado la clasificación de los ORI de metazoos en tres tipos diferentes: a) pobres en marcas epigenéticas, asociados a replicación tardía y poco eficientes b) correlacionados con marcas epigenéticas como H3K4me1 y H3K36me3 ligadas a regiones *enhancers* aunque con una eficiencia relativamente baja, y c) asociados a complejos Polycomb, cromatina abierta y altamente eficientes (Cayrou *et al.*, 2015).

* * *

Los datos de localización y actividad de los ORI en distintos sistemas coinciden en la existencia de una estrecha relación entre los procesos de replicación y transcripción. Dado que sendos procesos ocurren en el contexto de una estructura compleja de cromatina y que la iniciación de ambos se ha relacionado con la presencia de conformaciones del DNA no canónicas, como *R-loops* y estructuras de G4, resulta de interés analizar si estas correlaciones son más causales que casuales. En los trabajos que conforman esta Tesis Doctoral hemos querido profundizar especialmente en la relación entre la estructura de la cromatina y el inicio de la replicación del DNA

combinando técnicas genómicas con estudios detallados a alta resolución. Hemos utilizado distintos modelos celulares: células de mamífero (ratón y humano) y promastigotes de *Leishmania major* para aprovechar las distintas ventajas que ofrecen ambos sistemas en cuanto a los datos epigenéticos disponibles en la literatura y a la organización de sus genomas.

Los objetivos que establecimos para la realización de esta Tesis Doctoral se enmarcan en el propósito más amplio de la comprensión de la regulación y la dinámica alrededor de los ORI. Para ello establecimos como prioridades:

1. Estudiar a nivel genómico la estructura y composición de la cromatina junto con la abundancia de intermediarios de replicación del DNA en sistemas de células de mamíferos (ratón y humano), integrando además la información disponible respecto a la presencia de motivos de secuencia y localización de proteínas de replicación.
2. Analizar a alta resolución la estructura de la cromatina y la relación con la abundancia de intermediarios de replicación del DNA en ejemplos representativos de ORI bien caracterizados en células de ratón.
3. Generar el perfil nucleosomal y de iniciación de la replicación del DNA en un modelo con un genoma altamente compartimentalizado respecto a la actividad transcripcional; el protozoo *Leishmania major*.
4. Integrar los datos epigenéticos disponibles junto con el perfil nucleosomal, la abundancia de intermediarios de replicación del DNA y la cinética transcripcional para profundizar en la relación entre estos procesos nucleares en *L. major*.

1. Manipulación de material biológico

1.1 Células de ratón

El tipo celular utilizado en los estudios en células de ratón fueron líneas troncales embrionarias PGK12.1 (Penny *et al.*, 1996). Su manipulación se llevó a cabo en condiciones asépticas en campana de seguridad biológica de tipo P2 (Faster BH 2004) y manteniendo los cultivos en un incubador (Esco Celculture) a 37°C, 5% CO₂ y con humedad.

Se consideró el cultivo en fase exponencial al alcanzar una confluencia visual del 70-80% de crecimiento en los frascos de cultivo. Confluencias superiores al 90% se determinaron asociadas a la fase estacionaria. Las células PGK12.1 se cultivaron en frascos de cultivo T25 o T75 (Falcon) previamente gelatinizados utilizando como medio de cultivo DMEM sin piruvato de sodio (Gibco) suplementado con 20% suero fetal bovino (FBS) testado para células ES (Lonza), 1 mM piruvato de sodio (Gibco), 1x aminoácidos no esenciales (Gibco), 2 mM L-glutamina (Gibco), 50 μM β-mercaptoetanol (Gibco), 1:100 PenStrep (Gibco) -concentraciones finales de 100 U/mL penicilina y 100 μg/mL streptomycin- y 1000 U/mL factor inhibidor de leucemia (ESGRO-LIF, Millipore).

Una vez alcanzada la fase exponencial de crecimiento en el volumen determinado, las células se expandieron en una dilución 1:6 o 1:8. Para ello, se realizaron dos lavados con PBS a temperatura ambiente. Posteriormente se incubaron con 1-2 mL de 0,05% tripsina-EDTA (Gibco) suplementada con 2% suero de pollo hasta el total desprendimiento celular (2-5 minutos). Inmediatamente después, se diluyeron en medio de cultivo para inactivar la tripsina. Esta solución se dividió por el factor de dilución determinado y se trasvasó a nuevos frascos de cultivo gelatinizados completándolos con medio de cultivo hasta el volumen apropiado.

Para el almacenaje a largo plazo de viales celulares, alcanzada la confluencia del 70-80% se procedió de manera similar a lo descrito pero la tripsina se inactivó con medio de congelación -90% FBS y 10% DMSO (Sigma-Aldrich)- hasta una concentración aproximada de 1-3x10⁶ células/mL. Posteriormente, se prepararon alícuotas de 1 mL en crioviales y se procedió a su congelación gradual hasta -80°C en un recipiente conteniendo isopropanol durante 12h. Finalmente, los viales se transfirieron a nitrógeno líquido. Para la reactivación de las células desde su almacenamiento en nitrógeno líquido, se descongelaron los viales unos segundos a 37°C e inmediatamente se trasvasaron a un volumen de medio de cultivo al menos

diez veces mayor (para la dilución del DMSO). A continuación, se centrifugó 3-5 minutos a 1000-2000 rpm, se sustituyó el sobrenadante por medio de cultivo y se transfirió a frascos de cultivo previamente gelatinizados.

1.2 Células de *Leishmania major*

La cepa de parásito utilizada como modelo fue *Leishmania major* Friedlin. Los cultivos se manipularon en condiciones asépticas en campana de seguridad biológica de tipo P2 (Esco Labculture) y se mantuvieron en estufa (J.P. Selecta) a 26°C.

Para definir el crecimiento de los cultivos se tomaron como valores de referencia 6×10^6 parásitos/mL como fase exponencial y $>10^7$ parásitos/mL como fase estacionaria. Los parásitos se cultivaron en medio M199 (Sigma-Aldrich) suplementado con 10 % de suero fetal bovino inactivado, 40 mM HEPES (pH 7,4), 0,1 mM adenina, 10 µg/mL hemina, 1 µg/mL biotina, 2 ng/mL biopterina, 100 U/mL penicilina G y 0,1 mg/mL sulfato de estreptomicina.

2. Purificación de DNA genómico de alto peso molecular

2.1 Células de ratón

La obtención de DNA genómico se realizó mediante el procedimiento estándar de Sambrook (1989). Una vez que las células alcanzaron la fase estacionaria de crecimiento, se realizaron dos lavados con PBS y se añadió a cada frasco de cultivo 1-2 mL de solución de lisis -50 mM TRIS (pH 8), 10 mM NaCl, 10 mM EDTA (pH 8) y 0,5% SDS-. Se agruparon los lisados en un único tubo, se suplementaron con 100 µg/mL proteinasa K (Roche) y se dejaron en incubación 12 horas a 37°C. A continuación se mezcló el lisado suavemente en tubos de 50 mL (Falcon) con un volumen de fenol saturado con 10 mM TRIS (pH 8) y 1 mM EDTA- (Sigma) y se centrifugó 10 minutos a 3500 rpm. Se repitió el proceso combinando la fase acuosa con una mezcla 25:24:1 de fenol:cloroformo:alcohol isoamílico. Finalmente, el DNA se precipitó con 2 volúmenes de etanol absoluto, se lavó con 70% etanol, dejó secar y resuspendió en 1 mL TE -10 mM TRIS (pH 8), 1 mM EDTA- durante 24-48 horas a 4°C.

2.2 Células de *L. major*

Para la obtención del DNA genómico de los parásitos se recogieron los cultivos crecidos hasta fase exponencial en tubos de 50 mL (Falcon) mediante centrifugación durante 5 minutos a 2500 rpm a 4°C (Sorvall RT 6000B). Los precipitados celulares se lavaron dos veces con PBS frío y se purificaron los núcleos adaptando el protocolo de Gong *et al.* (1996).

Para la obtención de núcleos, se trabajó en todo momento con las soluciones en frío y manteniendo todo el material en hielo. Los precipitados celulares se resuspendieron a una concentración de $1,7 \times 10^7$ células/mL en tampón de homogeneización -10 mM TRIS (pH 7,4), 1 mM EDTA (pH 8), 0,1 mM EGTA (pH 8), 15 mM NaCl, 50 mM KCl, 0,2% NP-40 y 5% sacarosa suplementado con 0,15 mM espermina y 0,5 mM espermidina justo antes de su utilización-. Después de una incubación en hielo de al menos 5 minutos (para la ruptura de la membrana celular), se transfirieron gota a gota con pipeta de vidrio a tubos de 50 mL en los que se había añadido un colchón de sacarosa -tampón de homogeneización con 10% de sacarosa-. Se prepararon 2 tubos (de 50 mL) con 15 mL de colchón de sacarosa sobre los que se añadieron 22,5 mL de lisado y se centrifugaron 20 minutos a 2300 rpm a 4°C. Los sobrenadantes con los detritus celulares se descartaron y los precipitados conteniendo los núcleos purificados se resuspendieron y agruparon en un volumen final de 5 mL de tampón de lisis -50 mM TRIS (pH 8), 0,1 M EDTA (pH 8), 0,15 M NaCl, 0,5% SDS y 0,1 mg/mL proteinasa K- y se incubaron a 37°C durante 12 horas. La extracción y precipitación del DNA se llevaron a cabo de la misma forma que en el caso anterior (apartado 2.1).

3. Purificación de intermediarios de replicación de pequeño tamaño (SNS)

3.1 Células de ratón

Los intermediarios de replicación se fraccionaron por tamaño en gradientes de sacarosa (Gómez y Antequera, 2008). Para cada tubo de gradiente, se recogieron 4-5 frascos de cultivo T75 en fase exponencial de crecimiento y se extrajo el DNA genómico tal y como se describe en el apartado 2.1 pero suplementando los lisados con 0,1 U/ μ L RNaseOUT (Invitrogen).

Cada gradiente se preparó a partir de siete diluciones de sacarosa con concentraciones desde 5% a 20% en tampón TEN -10 mM TRIS (pH 8), 1 mM EDTA y

100 mM NaCl- cargadas cuidadosamente con bomba peristáltica en tubos de 13 mL de ultracentrífuga (Beckman Coulter). El DNA genómico homogéneamente resuspendido se desnaturalizó a 95°C durante 5-10 minutos y se añadió cuidadosamente en la parte superior del gradiente. A continuación, los tubos se centrifugaron en una ultracentrífuga Beckman Coulter Optima™ L-100 XP durante 20 horas a 24000 rpm y 20 °C utilizando un rotor SW40Ti. Después se recogieron 13 fracciones de 1 mL desde la parte superior del gradiente y se precipitaron en tubos de 1,5 mL (Eppendorf) con 1:10 volúmenes de 3 M acetato de sodio y 2 volúmenes de 100% etanol, se lavaron con 70% etanol y se resuspendieron en 60 µL TE.

Posteriormente se analizó el perfil del fraccionamiento mediante electroforesis alcalina (**Figura 5A**). Para ello, se desnaturalizó el 10% de cada fracción -con 0,2 M NaOH durante 15 minutos- y se examinó en geles de agarosa en condiciones alcalinas -50 mM NaOH, 1 mM EDTA, 1% agarosa-, realizándose la electroforesis en el mismo tampón (50 mM NaOH, 1 mM EDTA) a 4°C. Finalmente, el gel se neutralizó con 2x SSC dos veces durante 30 minutos y el DNA se visualizó mediante tinción con bromuro de etidio. Las fracciones del gradiente conteniendo los intermediarios de replicación más pequeños y, por tanto, más próximos al origen de replicación fueron las fracciones F3 (150-600 bp), F4 (300-800 bp) y F5 (400-1200 bp).

Después del fraccionamiento por tamaño, se procedió a un enriquecimiento bioquímico en intermediarios de replicación (protegidos en el extremo 5' por el cebador de RNA) del resto de fragmentos contaminantes, procedentes de rupturas del DNA, mediante tres rondas consecutivas de fosforilación y digestión con λ -exonucleasa. Tras una desnaturalización de 5 minutos a 95°C, se llevó a cabo la reacción de fosforilación en un volumen de 200 µL en presencia de 1 mM ATP (Fermentas), 0,5 U/µL PNK (Fermentas) y 0,2 U/µL RNaseOUT. Después de 30 minutos, las reacciones se pararon con 1:10 de volumen de solución STOP -1,25% sarkosyl, 125 mM EDTA y 310 µg/mL proteinasa K - durante 30 minutos a 50°C. A continuación, las muestras se extrajeron con fenol/cloroformo, precipitaron, lavaron y resuspendieron en 100 µL de agua. Posteriormente, se desnaturalizaron de nuevo 5 minutos a 95°C y se realizó el enriquecimiento bioquímico en un volumen de 200 µL con 0,75 U/µL λ -exonucleasa (Fermentas) en presencia de 0,2 U/µL RNaseOUT durante 12h a 37°C. Finalmente, las reacciones se pararon por incubación a 75°C durante 10 minutos y, de manera similar, se extrajeron con fenol/cloroformo, precipitaron y resuspendieron en 60 µL TE.

Paralelamente se realizaron controles para evaluar la digestión de fragmentos no protegidos por RNA en su extremo 5'. Para ello se añadieron 100 ng de un plásmido de 3 Kb linearizado y fosforilado en su extremo 5' a alícuotas con 1:10 de cada uno de los tubos que contenían las reacciones de λ -exonucleasa. Si la acción de esta última era efectiva, al revelar el experimento mediante electroforesis en gel de agarosa, desaparecería la banda correspondiente al plásmido utilizado como control (**Figura 5B**). De este modo, observábamos que cuando se incorporaban 15 U de λ -exonucleasa (el equivalente a 1:10 de la actividad presente en las digestiones de las muestras) el plásmido se digería completamente (carril 1); cuando no se añadía enzima (control negativo) el plásmido se mantenía íntegro con un tamaño de 3 Kb (carril 2) y cuando se aportaba el plásmido a alícuotas correspondientes a 1:10 de las distintas fracciones de intermediarios de replicación, era nuevamente degradado (carriles 3 y 4).

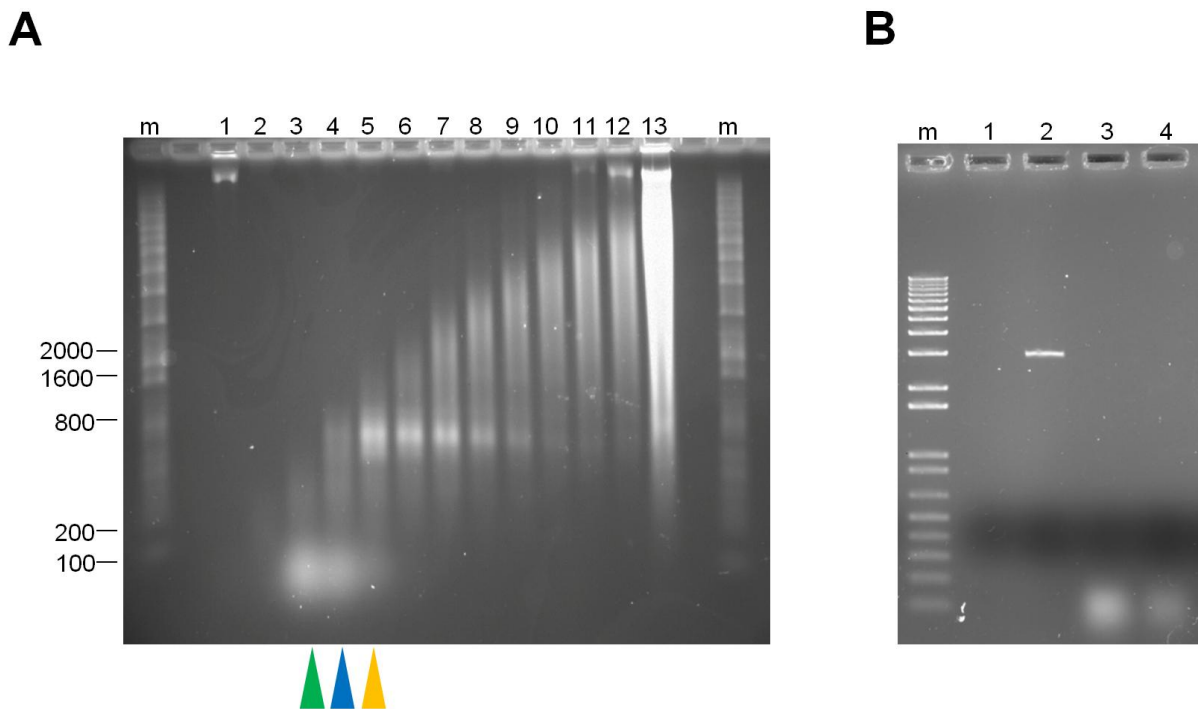


Figura 5: **A:** Ejemplo representativo del análisis del fraccionamiento del DNA genómico de *L. major* en gradientes de sacarosa. A la izquierda se indican los tamaños de los fragmentos de DNA conforme al marcador molecular (m) 1 Kb plus DNA Ladder (Invitrogen). Las fracciones señaladas con flechas de color en la parte inferior son las utilizadas para los diversos experimentos de este proyecto: F3 (150-600 bp), F4 (300-800 bp) y F5 (400-1200 bp). Nótese una banda que migra aproximadamente a 700-800 bp y que se corresponde con el kinetoplasto del parásito (esta banda no aparece en los modelos de vertebrados). **B:** Control de digestión de la enzima λ -exonucleasa que demuestra que el DNA lineal y fosforilado de 3 Kb utilizado como control (carril 2: ssDNA - λ exo) queda completamente digerido en estas condiciones por la enzima tanto si está sólo en la muestra (carril 1: ssDNA + λ exo) como acompañado por DNA proveniente de intermediarios de replicación (carriles 3 y 4: ssDNA + F3/F4 + λ exo). *m* es el mismo marcador molecular que en el caso anterior.

3.2 Células de *L. major*

Para cada gradiente de sacarosa, se recogieron 500 mL de cultivo de células en fase exponencial de crecimiento y se purificó el DNA genómico de la manera descrita en el apartado 2.2, con algunas modificaciones: se prepararon un total de 8 tubos de 50 mL, cada uno de ellos conteniendo 15 mL de colchón de sacarosa sobre los que se añadieron 22,5 mL de lisado celular y una vez obtenidos los núcleos, el lisado se suplementó con 0,1 U/ μ L RNaseOUT.

El enriquecimiento en intermediarios de replicación mediante digestión con λ -exonucleasa se llevó a cabo de manera similar a lo descrito en el modelo de ratón (apartado 3.1).

4. Purificación de DNA mononucleosomal

Para ambos modelos, el protocolo de preparación de DNA mononucleosomal (MN) fue precedido de la purificación de núcleos de la manera descrita en el apartado 3 con algunas modificaciones. En el caso de *L. major*, al ser necesarios volúmenes menores que para la obtención de SNS, los colchones de sacarosa para la purificación de núcleos se realizaron en tubos de 15 mL (Falcon), depositando 6 mL de tampón de homogeneización (10^8 células o parásitos) sobre 3,5 mL de colchón de sacarosa.

Después de la centrifugación a través de los colchones, se resuspendió el precipitado de núcleos en otros 6 mL de solución de lavado -10 mM TRIS (pH 7,4), 15 mM NaCl₂ 50 mM KCl y 8,5% de sacarosa suplementado con 0,15 mM espermina y 0,5 mM espermidina justo antes de su utilización- (Gong *et al.*, 1996). El contenido se dividió en seis alícuotas de 1 mL cada una (aproximadamente $1,7 \times 10^7$ células por alícuota), que serían digeridas por las cantidades correspondientes de enzima MNasa.

Las concentraciones de MNasa utilizadas por cada $1,7 \times 10^7$ núcleos fueron 300 U, 600 U, 1200 U y 2400 U, en el caso de las células de ratón, y 19,5 U, 37,5 U, 75 U, 150 U y 300 U, en el caso de células de *L. major*. En ambos casos, la digestión se realizó durante 6 minutos a 25°C, después de lo cual las reacciones se pararon añadiendo 9 mM EDTA y 3,5 mM EDTA y transfiriéndolas a hielo. Las alícuotas digeridas en las mismas condiciones se agruparon, se incubaron durante 1-2 horas a 37°C con RNase-DNase free (Roche) y posteriormente en solución de lisis durante 12h. Por último, el DNA resultante se purificó y resuspendió en 50-100 μ L TE.

El grado de digestión alcanzado con concentraciones crecientes de MNasa se analizó mediante electroforesis en geles de agarosa al 2,5%, observándose un patrón en escalera característico en función de la intensidad de la digestión (**Figura 6A**). En el caso de los experimentos de escaneo de DNA mononucleosomal mediante amplicones solapantes (NuSA), como los realizados en ORI concretos en células de ratón, se purificaron las bandas de 140-150 bp de cada una de las digestiones realizadas. En el caso de los experimentos de secuenciación masiva de DNA mononucleosomal (MNasa-seq) de *L. major*, se eligió la digestión de 150 U que correspondía aproximadamente a una relación 80:20 en mono:dinucleosomas. Las bandas escindidas de los geles de agarosa se purificaron mediante el kit Wizard® SV Gel and PCR Clean-Up System (Promega).

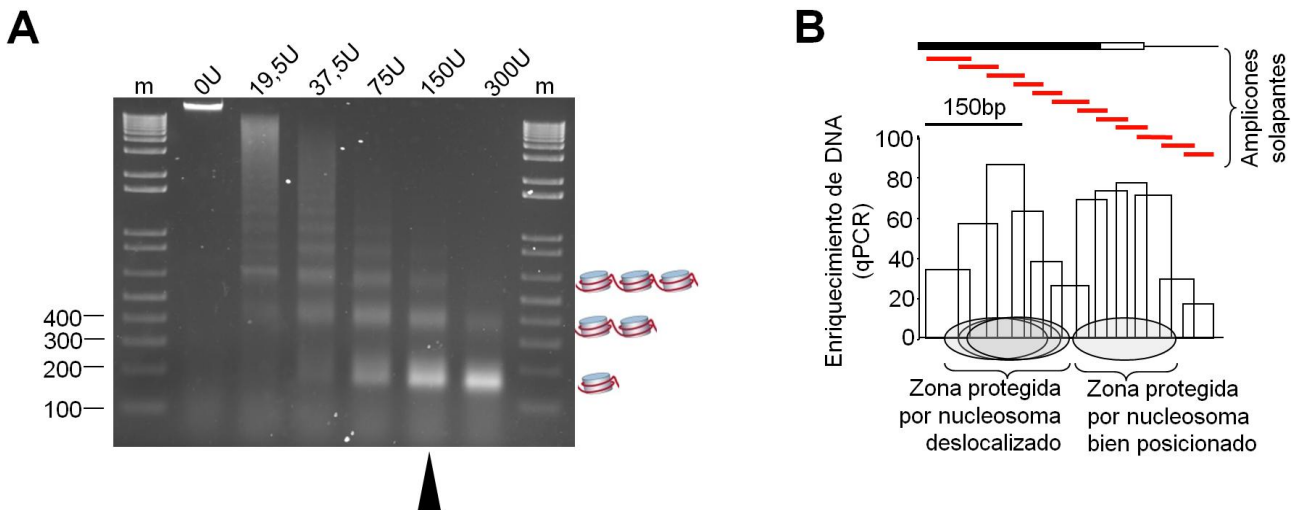


Figura 6: **A:** Ejemplo representativo del análisis de la fragmentación del DNA de *L. major* generado tras la digestión con cantidades crecientes de MNasa (en células de ratón se requieren mayores cantidades de enzima para alcanzar fragmentaciones equivalentes). La muestra señalada contiene una proporción 80:20 en mono:dinucleosomas, es la utilizada en los experimentos de MNasa-seq. A la izquierda se indican los tamaños de los fragmentos de DNA conforme al marcador molecular (m) 1 Kb plus DNA Ladder (Invitrogen); a la derecha, la posición correspondiente al DNA proveniente de mono, di o tri-nucleosomas. **B:** Esquema ilustrativo del perfil de ocupación nucleosomal obtenido mediante experimentos de NuSA. La cromatina más expuesta es digerida por la enzima presentando forma de “valle” mientras que las regiones protegidas por nucleosomas forman un “pico” que se extiende 150 bp o más, en función del nivel de posicionamiento del nucleosoma en la población celular analizada.

4.1 Ensayo de escaneo de nucleosomas (NuSA)

Este ensayo (Sekinger *et al.*, 2005) consiste en la cuantificación mediante PCR en tiempo real (qPCR) de la abundancia de amplicones de 60-80 bp solapantes entre sí en 15-25 bp en preparaciones de DNA mononucleosomal provenientes de exposiciones crecientes a MNasa. De esta manera, si los amplicones están situados en zonas expuestas de la cromatina -por tanto, digeridas por la MNasa- se obtiene un bajo enriquecimiento, mientras que éste es alto en las regiones protegidas por nucleosomas (más resistentes a la digestión por MNasa). Típicamente se detecta un patrón de picos y valles correspondientes a aquellas regiones respectivamente más o menos ocupadas por nucleosomas a lo largo de la población celular (**Figura 6B**).

5. PCR cuantitativa (qPCR)

Las reacciones de PCR cuantitativa a tiempo real (qPCR) se realizaron en placas de 384 pocillos (Applied Biosystems) en un instrumento ABI Prism 7900HT (Applied Biosystems). Las reacciones se prepararon en 20 μ L de volumen final con 0,025 U/mL HotStar Taq Polimerasa (Qiagen) y sus reactivos -1x tampón y 1,5 mM $MgCl_2$ - suplementados con 0,2 mM dNTPs (Roche), 375 nM oligonucleótidos (Sigma-Aldrich) y 1x SYBR Green (Molecular Probes). Las secuencia de los oligonucleótidos utilizados y las condiciones de reacción de cada amplicón se indican en **Anexo I: Tabla de oligonucleótidos utilizados**.

Para la definición de las curvas de amplificación, por cada uno de los amplicones analizados se utilizaron cinco diluciones seriadas 1:5 de DNA genómico sonificado en la misma mezcla de reacción que las muestras experimentales. Las mejores condiciones para cada pareja de oligonucleótidos se determinaron empíricamente, aceptándose aquellas con una eficiencia de reacción de $100\% \pm 10$ -es decir, una pendiente de la curva de reacción de $3,3 \pm 0,3$ - y un coeficiente de correlación entre los puntos de $R^2 > 0,99$.

Los análisis de los resultados de qPCR se realizaron con el programa ABI Prism 7900HT SDS Software (versión 2.4).

6. Secuenciación masiva (NGS)

6.1 Preparación de librerías de SNS-seq y MN-seq de *L. major*

Como paso previo a generar las librerías de SNS, los intermediarios de replicación provenientes de las fracciones F4 y F5, compuestas por DNA de cadena sencilla, se transformaron en DNA de cadena doble (Cadoret *et al.*, 2008). Tras la última ronda de tratamiento con λ -exonucleasa, las muestras se resuspendieron en un volumen de 29 μ L a los que se añadieron 1 μ L (50 pmol) de hexámeros al azar fosforilados (Roche) y se incubaron durante 5 minutos a 95°C permitiendo el anillamiento a las SNS mediante enfriamiento gradual hasta alcanzar los 4°C. A continuación, se realizó la extensión de los oligonucleótidos en presencia de 1 μ L (10 mM) dNTP (Roche) y 1 μ L (5 U) *exo*-Klenow (New England Biolabs) durante 1 hora a 37°C y se detuvo la reacción por incubación 10 minutos a 75°C. Posteriormente los fragmentos de las cadenas complementarias se incubaron con 2 μ L (80 U) de Taq DNA ligasa (New England Biolabs) durante 30 minutos a 50°C y se detuvo la reacción por incubación 10 minutos a 75°C. Finalmente, se eliminaron los cebadores de RNA de las SNS por digestión con 1 μ L de mezcla de RNAsas A/T1 (Thermo Scientific) durante 30 minutos a 37°C, se realizó una extracción con fenol/cloroformo, precipitación y resuspensión del DNA en 20 μ L de TE. Como control se preparó una librería de DNA genómico proveniente de células en fase estacionaria (10^9 parásitos/mL) fragmentado por sonicación con 30 ciclos de 30 segundos ON/OFF en Bioruptor® (Diagenode) hasta alcanzar un tamaño de 200-500 bp.

El DNA mononucleosomal proveniente de la digestión con MNasa ya era de doble cadena y no requería fragmentación por lo que no demandaba modificaciones previas en la generación de las librerías.

Las librerías de secuenciación se generaron en la unidad de Genómica del Parque Científico de Madrid (PCM) mediante el protocolo estándar de construcción de librerías TruSeq (Illumina) con algunas modificaciones. Se evitó la fragmentación de las muestras porque ya había sido realizada durante los distintos protocolos de purificación. Las librerías de SNS se purificaron posteriormente en geles de poliacrilamida para eliminar la contaminación con dímeros de adaptadores vacíos, evitando así el paso de purificación habitual en columnas de Ampure. Para minimizar la introducción de artefactos en la posterior secuenciación, se limitaron al mínimo los ciclos de PCR (10-15 ciclos) en la generación de las librerías.

6.2 Secuenciación

La secuenciación de las librerías de SNS y MN se realizó en el Centro de Regulación Genómica (CRG-Barcelona) y en el Centro Nacional de Análisis Genómico (CNAG-Barcelona), primera y segunda réplica respectivamente. En todos los casos los fragmentos de las librerías se secuenciaron por ambos extremos (*paired-end*) con lecturas de 50 bp (2x50) en un equipo HiSeq 2000/2500 (Illumina).

7. Análisis bioinformáticos

7.1 Análisis de datos de NGS de repositorios públicos provenientes de células de ratón y humano

La versión utilizada del genoma de ratón fue mm9 y la del genoma humano hg18. Cuando los datos utilizados se habían generado en versiones anteriores de los respectivos genomas, los alineamientos se convirtieron a las versiones actuales utilizando la herramienta *LiftOver* del navegador genómico de la Universidad de California en Santa Cruz (*UCSC Genome Browser*) (Kent *et al.*, 2002).

7.1.1 Selección de los sitios de inicio de la transcripción (TSS)

Los TSS de ambos genomas se definieron a partir del extremo 5' de los transcritos mayores de 200 bp (para evitar pseudogenes o genes codificados por otras polimerasas) de la base de datos RefSeq (Pruitt *et al.*, 2005).

7.1.2 Utilización de datos accesibles de células de ratón

Los datos del mapeo de intermediarios de replicación y localización de ORI derivados de estos se obtuvieron a partir de trabajos publicados de hibridación de SNS en *microarrays* genómicos (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011). Las regiones definidas en el primero de ellos como ORI se convirtieron desde mm7 a mm9 y las densidades de SNS se generaron a partir de los perfiles de *microarrays*. Las densidades de SNS del segundo estudio se transformaron desde mm8 a mm9 y los ORI se definieron como regiones de menos de 5 Kb con al menos 10 sondas consecutivas positivas cuyo su valor de intensidad promedio fuese superior a 0,5.

Los perfiles de nucleosomas se obtuvieron a partir de la transformación de lecturas de NGS (Teif *et al.*, 2012) en los correspondientes perfiles de densidad.

7.1.3 Utilización de datos accesibles de células humanas

El paisaje de iniciación de la replicación se generó a partir de las lecturas de SNS de células HeLa (Besnard *et al.*, 2012) y para la definición de ORI se utilizó el algoritmo de MACS2 versión 2.1.0.20140616 (Zhang *et al.*, 2008) en modo *callpeak* evitando la generación de modelo, instando una extensión de las lecturas de 1500 bp y limitando el valor de falsos positivos al 1% (`--nomodel --extsize 1500 --q 0.01`).

Los perfiles de densidad de las variantes de histonas H3.3 y H2A.Z se generaron a partir de los datos de ChIP-seq publicados por Jin *et al.*, (2009).

Los datos de densidad y regiones de enriquecimiento significativo en ORC1 se obtuvieron a partir de los datos de ChIP-seq publicados en Dellino *et al.*, (2013).

Para seleccionar aquellos intervalos con presencia de *R-loops*, se trabajó con las lecturas de DRIP-seq de forma similar al estudio original (Ginno *et al.*, 2012) con alguna modificación. Las lecturas se alinearon con BWA (Li y Durbin, 2009) y se asignó el número de ellas en cada intervalo delimitado por las enzimas de restricción utilizadas. Después se seleccionaron los intervalos que presentaban lecturas en los extremos y un enriquecimiento al menos cinco veces superior a la media genómica. Finalmente, se cruzaron con aquellos presentes en zonas de unión de ORC1.

Para realizar el perfil de zonas con asimetría en GC (*GC-skew*) se utilizaron las regiones así definidas en el estudio anterior (Ginno *et al.*, 2012).

Los motivos de cuádruplex de guanina (G4) se identificaron con el programa Quadparser (Huppert y Balasubramanian, 2005), seleccionando los que respondían a una estructura $G_3-L_{1-7}-G_3-L_{1-7}-G_3-L_{1-7}-G_3$ (siendo G la base guanina y L cualquiera de las cuatro bases) (parámetros *GCgc 3 4 1 7*).

7.1.4 Manejo y tratamiento de los datos

Los datos se visualizaron en el navegador genómico de la Universidad de California en Santa Cruz (*UCSC Genome Browser*) (Kent *et al.*, 2002).

Las conversiones entre formatos BAM y SAM, ordenamiento de lecturas, creación de índice, selección de características del alineamiento, colapso de lecturas repetidas, etc. se llevó a cabo con el paquete SamTools (Li *et al.*, 2009).

Las intersecciones (*intersectBed*), colapso (*mergeBed*) y ordenamiento (*sortBed*) de intervalos, transformación de lecturas en perfiles de densidad (*genomeCoverageBed*), conversión de ficheros (*bedGraphToBigWig*), etc. se realizaron con el paquete BedTools versión 2.7.0 (Quinlan *et al.*, 2010).

Para el manejo de ficheros, datos, etc. se utilizaron programas propios escritos en lenguajes PERL y BASH (*shell scripts*).

Las representaciones de datos en perfiles agregados se realizaron mediante un programa creado por el Dr. Bastolla (Unidad de Bioinformática, CBMSO) normalizando los valores totales en unidades tipificadas o *Z-score* [$Z=(x-\mu)/\sigma$].

7.2 Análisis de datos de NGS de *L. major*

7.2.1 Anotación de características genómicas a partir de datos públicos

El genoma de referencia de *L. major* que se utilizó (versión 6.0) se descargó de la base de datos de tripanosomátidos (TriTrypDB; Aslett *et al.*, 2010). Las coordenadas de las secuencias codificantes (CDS) se obtuvieron de la base de datos GeneDB (Logan-Klumpler *et al.*, 2012).

Los perfiles de modificación de histona H3 (AcH3) se crearon realineando los datos publicados (Thomas *et al.*, 2009) respecto al genoma de referencia mediante Bowtie2 versión 2.2.0 (Langmead *et al.*, 2012) y suavizados utilizando R versión 3.1.0 (R Core Team, 2014).

Los perfiles de enriquecimiento en base J se llevaron a cabo realizando un *peakcalling* de los datos publicados (van Leunen *et al.*, 2012) a través de MACS versión 1.4.2 (Zhang *et al.*, 2008) utilizando el modo *sharppeaks* y generando un modelo basado en picos de 200 bp de extensión aproximada (*--bw 200*).

La localización de las regiones relativas a la transcripción (coordenadas de TSS, TTS, PTU, pre-RNAm, IR, SSR y HT) se definieron a partir de los datos publicados (Rastrojo *et al.*, 2013), cruzando con las anotaciones descritas de base J y

Ach3. Del mismo modo, los límites de las regiones UTR se concretaron mediante la intersección de los datos de transcripción con las secuencias de los CDS.

La estabilidad de los pre-RNA_m se definió a partir de los datos de RPKM de Rastrojo *et al.* (2013). Estos pre-RNA_m se asignaron a baja, media o alta estabilidad dependiendo de si $\log(\text{RPKM})$ se encontraba entre los valores de 0-2 (1,8% de pre-RNA_m), 2-4 (87,6% de pre-RNA_m), o mayores de 4 (10,6% de pre-RNA_m) respectivamente.

Las regiones mal anotadas del genoma (zonas de sombra) se definieron a partir de aumentos o disminuciones abruptos en los perfiles de alineamiento del DNA genómico que habitualmente iban acompañados por artefactos en algunos de los datos de NGS restantes y se eliminaron de los análisis posteriores. También se eliminaron aquellas regiones que presentaban repeticiones, secuencias duplicadas, baja mapeabilidad, etc. La localización de estas zonas de sombra se recoge en **Anexo I: Zonas excluidas del análisis.**

Los motivos G4 se obtuvieron con el programa Quadparser (Huppert y Balasubramanian, 2005) seleccionando los que respondían a una estructura $G_3-L_{1-7}-G_3-L_{1-7}-G_3$ (siendo G la base guanina y L cualquiera de las cuatro bases) (parámetros *GCgc 3 4 1 7*).

7.2.2 Alineamiento de lecturas de NGS

Para los alineamientos de las lecturas se utilizaron los alineadores Bowtie versión 1.0.0 (Langmead *et al.*, 2009) para los experimentos de SNS-seq y BWA -dentro del grupo de programas RUBioSeq (Rubio-Camarillo *et al.*, 2013)- para los experimentos de MNasa-seq. Cuando se utilizó Bowtie se hizo permitiendo hasta tres bases no coincidentes a lo largo de toda la lectura (*-v3*) y aceptando el mejor de hasta un máximo de cinco posibles alineamientos (*-m5 -k1 -best*). En el caso de BWA del paquete RUBioSeq se utilizó una modificación del protocolo que permitía hasta cuatro bases no coincidentes sin alterar el resto de parámetros. Para descartar aquellas lecturas que contenían motivos G4 en sus extremos 5', se realizó una intersección de las primeras 10 bp de sendos extremos frente a los motivos G4 y dichas lecturas se suprimieron del análisis -al haber convertido las SNS en dsDNA, no puede conocerse el extremo 5' del fragmento original, por lo que decidimos ser conservadores y considerar ambos como posibles-.

7.2.3 Definición de regiones significativas y construcción de perfiles

Una vez alineadas las lecturas de SNS-seq se procedió a colapsar las lecturas repetidas y a caracterizar cuáles eran los enriquecimientos significativos respecto a un control de DNA genómico -proceso de *peak calling*-. Para ello se utilizaron los programas MACS2 versión 2.1.0.20140616 (Zhang *et al.*, 2008) y SNS-scan (Picard *et al.*, 2014). MACS2 se utilizó en modo *callpeak* evitando la generación de modelo e instando una extensión de las lecturas de 150 bp (*--nomodel --extsize 150*). La definición de una región como ORI se hizo por extensión de 400 bp (para la fracción de 300-800 bp) o 600 bp (para la de 400-1200 bp) hacia ambos lados desde el punto máximo de enriquecimiento en SNS y los ORI solapantes se colapsaron considerándolos como un único origen. Los ORI comunes a las distintas réplicas biológicas se definieron como aquellos que aparecían en la fracción del gradiente que contenía intermediarios de replicación de menor tamaño (F4) y solapaban con los definidos en la fracción siguiente (F5). Para estimar la eficiencia de activación de cada ORI en la población de células se calculó el cociente entre el número de lecturas que definían el ORI por la extensión del mismo en bp. Para calcular la significación de la distribución genómica de los ORI caracterizados, se computó un reparto aleatorio por el genoma de intervalos del mismo tamaño que los ORI, repitiendo esta operación durante 10000 ciclos y comparando el p-valor de la distribución resultante respecto a la observada. La densidad de ORI se estableció como el número de bp de regiones definidas como ORI en ventanas de 20000 bp.

Para establecer el mapa de ocupación nucleosomal, en primer lugar se seleccionaron los pares con alineamiento de ambas lecturas en sentidos opuestos y una distancia coherente. Posteriormente se aplicó el software NUCwave (Quintales *et al.*, 2015) ajustando el perfil de alineamiento de las lecturas a una función de onda. Para la normalización en *Z-score* [$Z=(x-\mu)/\sigma$] se realizó previamente una corrección a través de la media de lecturas por nucleótido en ventanas de 100 Kb -evitando el sesgo de mayor profundidad de lecturas que mostraban las zonas de replicación temprana-. Las regiones con *Z-score* menor de -1 ($Z<-1$) se consideraron de baja ocupación nucleosomal. La densidad de regiones de baja ocupación nucleosomal se estableció como el número de bp de regiones definidas como de baja ocupación nucleosomal en ventanas de 10000 bp.

Los análisis de replicación temprana/tardía se realizaron a partir de los datos de Análisis de Frecuencia de Marcadores o MFA (Marques *et al.*, 2015), bien reproduciendo su metodología -agrupando las lecturas de las fases S-temprana, S-

tardía y G2 en segmentos no solapantes de 2500 bp- o bien incrementando la resolución a una escala comparable a la de los ORI detectados utilizando ventanas de 1000 bp solapantes en 10 bp a lo largo de cada cromosoma. Las regiones de replicación temprana se definieron como aquellas de más de 500 bp con un valor de MFA superior a 1,4.

7.2.4 Manejo y tratamiento de los datos

Todos los datos se cargaron y visualizaron en el navegador *Integrated Genome Browser* (IGV) versión 2.3.34 (Robinson *et al.*, 2011).

Al igual que en los análisis realizados en metazoos, gran parte del manejo de datos y ficheros se realizó con SamTools (Li *et al.*, 2009), BedTools (Quinlan *et al.*, 2010) así como con programas propios escritos en lenguajes PERL y BASH.

Los gráficos de visualización de datos de eficiencia de ORI, *heat maps* así como los perfiles agregados se prepararon con R versión 3.1.0 (R Core Team, 2014).



RESULTADOS

PARTE I

1. Estudio genómico de la estructura de la cromatina y la iniciación de la replicación en células de ratón

Para determinar la relación entre el paisaje nucleosomal y los ORI en células de ratón, utilizamos datos genómicos disponibles en la literatura, tanto de densidad de nucleosomas (Teif *et al.*, 2012), como de intermediarios de replicación de pequeño tamaño provenientes de un trabajo previo de nuestro laboratorio (Sequeira-Mendes *et al.*, 2009). En este estudio se habían identificado los ORI mediante la hibridación de SNS derivados de células pluripotentes (ES) en *arrays* genómicos (SNS-chip) que contenían un total de 10,1 Mb de los cromosomas 3 y X de ratón -aproximadamente el 0,4% del genoma- representado por sondas de 60 bp a una resolución promedio de 250 bp. Del repertorio de ORI identificados en dicho estudio elegimos los 36 asociados a CGI por las evidencias que indican que este grupo de ORI se activan con mayor eficiencia en células de mamíferos, tanto de ratón como humanas (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011; Besnard *et al.*, 2012). A pesar del reducido tamaño de la muestra, los análisis de perfiles agregados centrados en el TSS indicaron que este se encuentra embebido en una región de baja ocupación nucleosomal (NDR) definida por un nucleosoma bien posicionado (-1) en el extremo 5' y, en menor medida, en el extremo 3'. (**Figura 7A**). La abundancia de intermediarios de replicación en estas regiones presenta un perfil bimodal, a ambos lados del TSS, con los puntos máximos localizados en regiones en las que se ha superado el mínimo de ocupación nucleosomal.

Debido a la baja cobertura de las sondas en estos *arrays*, así como al limitado número de ORI definidos en este trabajo, decidimos extender el análisis a otro estudio más reciente de SNS-chip que cubría 60,4 Mb -aproximadamente el 2,4% del genoma de ratón- a una resolución de 100 bp (Cayrou *et al.*, 2011). Los perfiles agregados mostraron unos resultados comparables a los del análisis anterior: una menor ocupación nucleosomal en la zona del TSS -más dramática en el extremo 5' y más gradual en el 3'- y una distribución bimodal de cadenas nacientes a ambos lados del TSS, con los picos de mayor abundancia coincidiendo con regiones en las que se localizan nucleosomas (**Figura 7B**). En conjunto, estos resultados indican que la iniciación de la síntesis de DNA en los promotores asociados a islas CGI ocurre en regiones con alta ocupación nucleosomal.

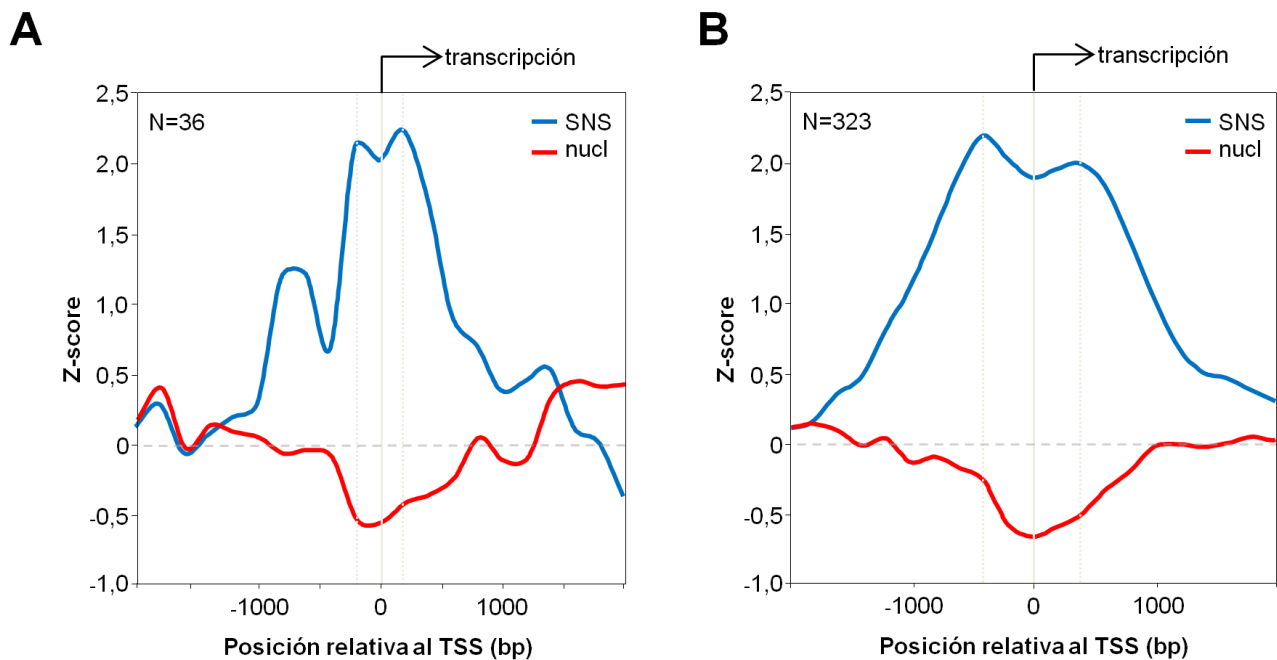


Figura 7: Distribución genómica de SNS y nucleosomas en ORI asociados a CGI en células ES de ratón. Los gráficos representan el Z-score de los valores de densidad nucleosomal (Teif *et al.*, 2012) y de densidad de intermediarios pequeños de replicación en ventanas de 200 bp a lo largo de 4 Kb de un total de **(A)** 36 ORI asociados a islas CpG (Sequeira-Mendes *et al.*, 2009) o de **(B)** 323 ORI asociados a islas CpG (Cayrou *et al.*, 2011). En el perfil agregado de nucleosomas (rojo) observamos una región de baja ocupación nucleosomal (NDR) en torno al TSS con un nucleosoma (-1) bien posicionado en dirección 5' desde el TSS. En el perfil agregado de SNS (azul) los máximos (líneas grises verticales) se sitúan en regiones de mayor abundancia nucleosomal.

2. Análisis local de alta resolución de la estructura de la cromatina y la iniciación de la replicación en células de ratón

Los resultados de los perfiles agregados no permitían analizar en detalle la relación entre la iniciación de la replicación y los nucleosomas individuales. Además, cada partícula nucleosomal puede presentar diferentes estados y propiedades físicas en distintas regiones de la cromatina (Tolstorukov *et al.*, 2010). Para abordar el estudio detallado de los nucleosomas de distintas regiones genómicas realizamos un análisis de alta resolución o escaneo de nucleosomas utilizando parejas de oligonucleótidos solapantes y qPCR sobre muestras de cromatina preparadas en diferentes condiciones (NuSA) (Sekinger *et al.*, 2005). Para inferir del modo más resolutivo posible la dinámica de los nucleosomas era necesario combinar distintos métodos de purificación de cromatina con un adecuado diseño de parejas de oligonucleótidos y condiciones de qPCR. Para cubrir un amplio intervalo de estados de cromatina, los núcleos celulares provenientes de cultivos asincrónicos en fase exponencial de

crecimiento previamente expuestos o no a formaldehído (*crosslinking*), se digirieron con cantidades crecientes de la enzima MNasa (que actúa sobre las regiones del DNA no protegidas por la unión de algún complejo proteico). Por otro lado, diseñamos amplicones de 60-80 bp solapantes en 15-25 bp de manera que pudieran generar un paisaje de picos -regiones resistentes a MNasa- y valles -regiones digeridas por MNasa- que reflejara los distintos grados de labilidad de los nucleosomas según las condiciones de extracción de la cromatina (Ver Materiales y Métodos). Para comprobar que la protección del DNA era efectivamente consecuencia de la presencia de un nucleosoma, llevamos a cabo un escaneo similar sobre cromatina tratada con formaldehído e inmunoprecipitada con anticuerpos específicos frente a la histona H3 en condiciones de alta o baja concentración de sal.

Para estos análisis de alta resolución elegimos cinco ejemplos de ORI caracterizados en trabajos previos del laboratorio (Sequeira-Mendes *et al.*, 2009 y **Figura 8**) y escaneamos una región de aproximadamente 600 bp conteniendo de 2 a 4 nucleosomas. Tres de los ORI seleccionados están asociados a promotores de CGI -*Mecp2*, *Atp2b3* y *Hs6st2*- y dos a exones de genes -*Pias3* y *Syde1*-. Además, atendiendo a su enriquecimiento relativo analizado por qPCR sobre preparaciones de SNS de tamaño creciente, podíamos dividirlos en dos grupos: ORI más eficientes, es decir, que se activan en un mayor número de células de la población -*Mecp2* y *Atp2b3*- y ORI menos eficientes -*Hs6st2*, *Pias3* y *Syde1*-.

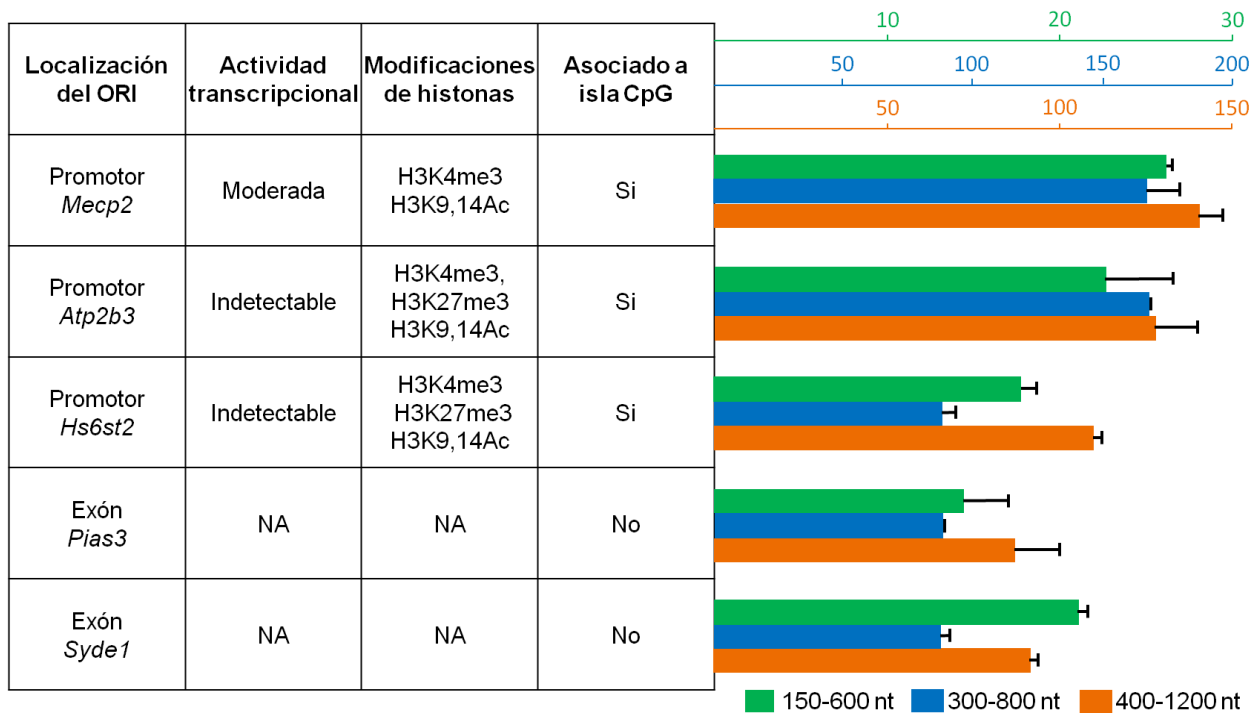
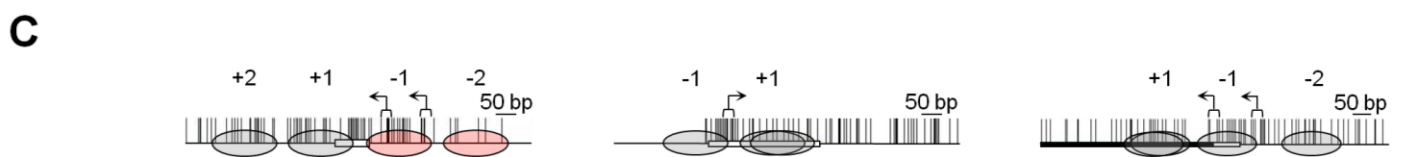
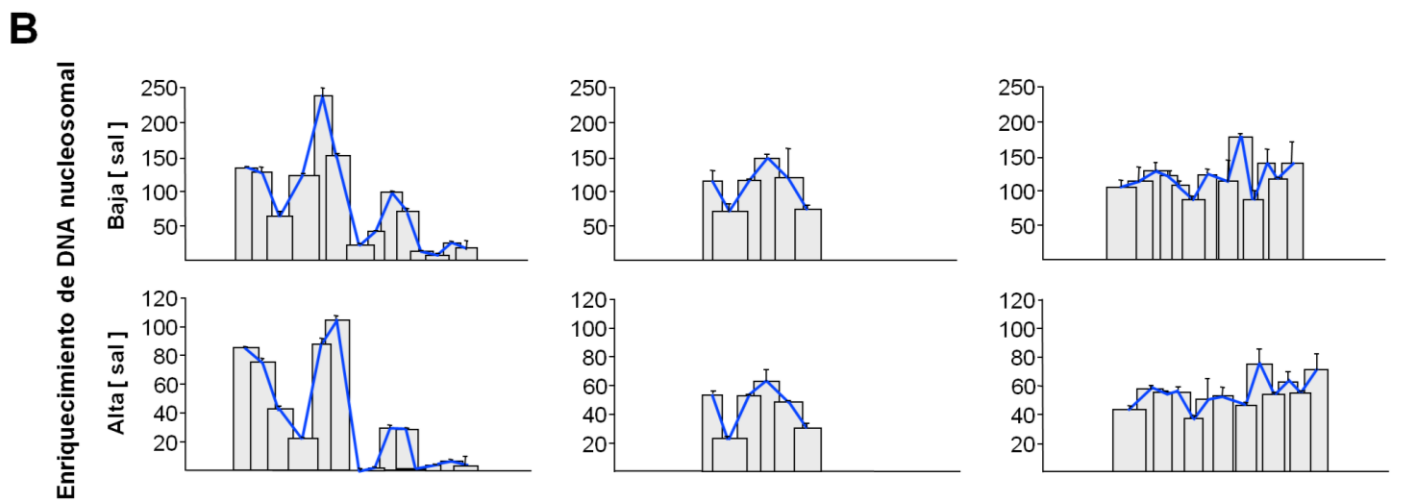
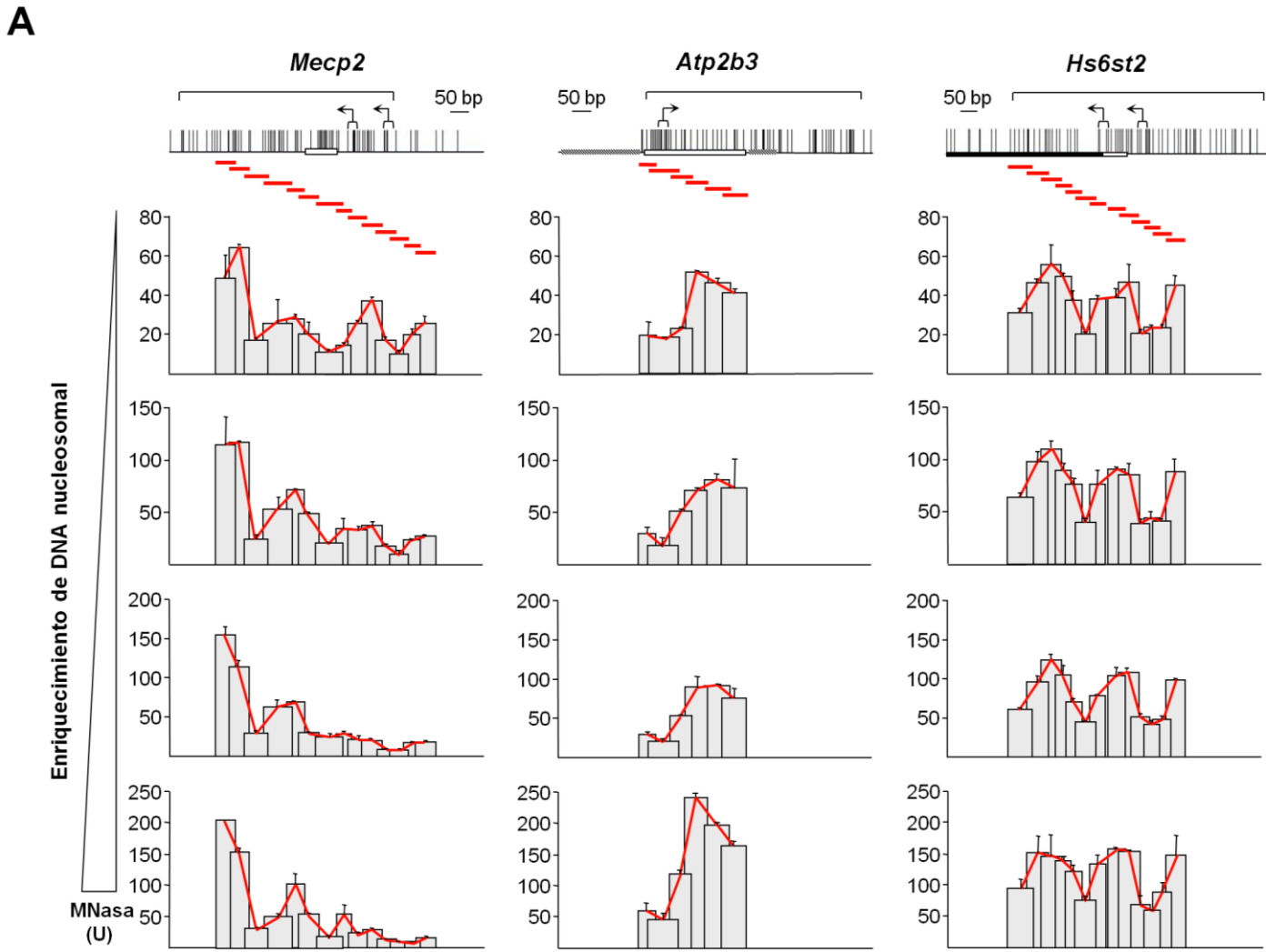


Figura 8: Características de los ORI seleccionados para este estudio. A la derecha, eficiencia relativa de los ORI en preparaciones de cadenas nacientes de los tamaños indicados. Los valores se normalizaron frente a los obtenidos para una región negativa en cada preparación de SNS.

En primer lugar, llevamos a cabo el análisis del posicionamiento y ocupación nucleosomal en los tres ORI asociados a promotores de CGI. En la **Figura 9** se representan los resultados del análisis de NuSA empleando cantidades crecientes de MNasa sobre cromatina nativa (panel A) o expuesta a formaldehído e inmunoprecipitada con anticuerpos anti-H3 en condiciones de baja o alta sal (panel B). En el caso del promotor-ORI del gen *Mecp2*, por ejemplo, identificamos cuatro nucleosomas bien posicionados, es decir, que protegen la misma región genómica de 150 bp en la mayoría de las células de la población (panel C): los situados en las posiciones +1 y +2 relativas al TSS los definimos como nucleosomas estables -protegen al DNA de la digestión por MNasa en todas las condiciones ensayadas-, mientras que solo detectamos los nucleosomas -1 y -2 en la digestión más suave y, por tanto, los definimos como nucleosomas lábiles. Los nucleosomas lábiles, o las regiones del DNA expuestas de manera transitoria, reflejan la existencia de un proceso continuo de remodelación de los nucleosomas necesario para mantener la accesibilidad de los elementos regulatorios (Deal y Henikoff, 2010). En el caso de la región promotora de *Mecp2*, la presencia de nucleosomas lábiles correspondía con la actividad transcripcional de este promotor (**Figura 8**). Por el contrario, en el caso de los promotores-ORI de los genes *Atp2b3*, y *Hs6st2*, cuya actividad transcripcional había sido indetectable en las células ES (**Figura 8**), observamos un paisaje nucleosomal invariable en todas las condiciones experimentales analizadas, lo que indicaba que estas regiones estaban empaquetadas con nucleosomas estables y bien posicionados (**Figura 9**).

Los nucleosomas de regiones no promotoras no suelen estar bien posicionados en la población celular (Valouev *et al.*, 2011). De manera consistente a estas observaciones, en los ORI asociados a exones de los genes *Pias3* y *Syde1* encontramos un nivel constante de densidad nucleosomal a lo largo de toda la región analizada, sin diferencias claras de enriquecimiento entre picos y valles (**Figura 10**). Estos resultados indican que estos dos ORI estaban empaquetados por nucleosomas estables pero que tenían un posicionamiento variable en las células de la población.

Figura 9: Análisis de NuSA mediante amplicones solapantes sobre tres ORI asociados a promotores. **A:** NuSA sobre cromatina preparada en condiciones nativas digerida con cantidades crecientes de enzima MNasa. En rojo, los amplicones utilizados en el análisis de qPCR. **B:** NuSA sobre cromatina tratada con formaldehído (*crosslinking*) e inmunoprecipitada con anticuerpos frente a la histona H3 en condiciones de baja (panel superior) y alta (panel inferior) concentración de sal. **C:** Interpretación del resultado y correspondencia con partículas nucleosomales: en gris si su estabilidad era alta (no se detectan cambios en las distintas condiciones empleadas) y en rojo si era baja (la protección a la MNasa se pierde a condiciones de digestión más estrictas).



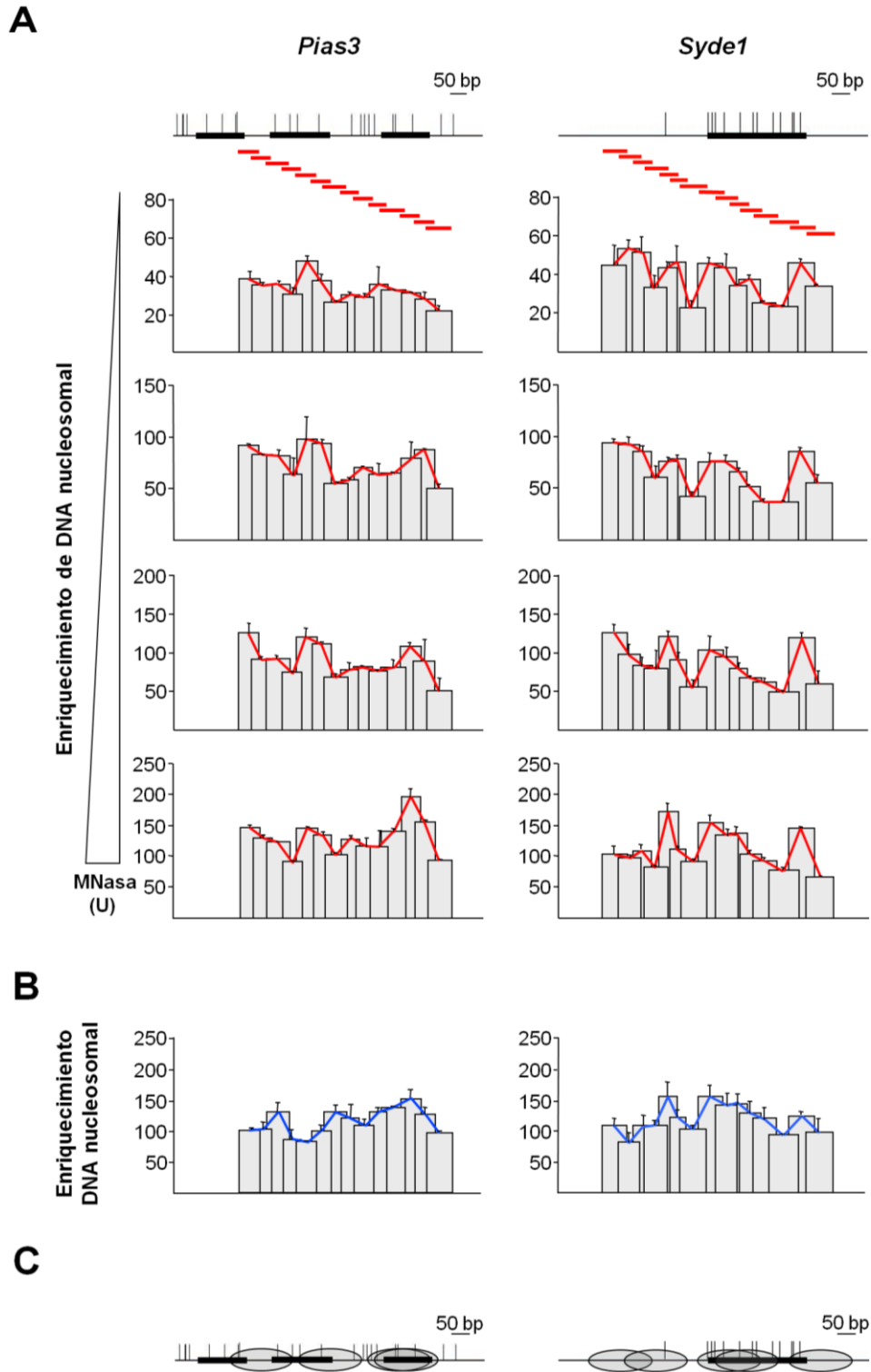


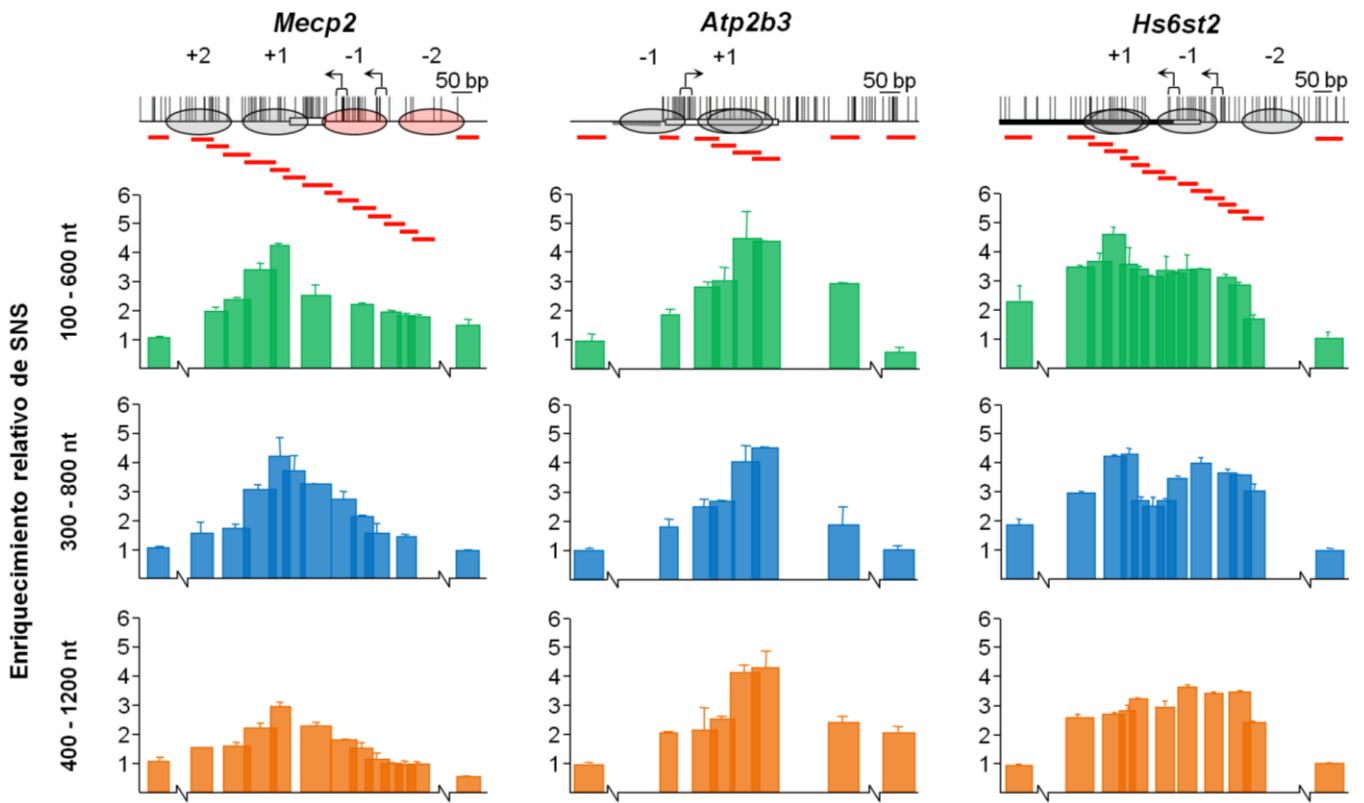
Figura 10: Análisis de NuSA por medio de amplicones de qPCR solapantes sobre dos ORI no asociados a promotores. **A:** NuSA sobre cromatina preparada en condiciones nativas digerida con cantidades crecientes de enzima MNasa. En rojo, los amplicones utilizados en el análisis de qPCR. **B:** NuSA sobre cromatina tratada con formaldehído (*crosslinking*) e inmunoprecipitada con anticuerpos frente a la histona H3 en condiciones de baja concentración de sal. **C:** Interpretación del resultado y correspondencia con partículas nucleosomales (símbolos con el mismo significado que en **Figura 9**).

Para realizar el análisis de iniciación de la replicación en estos ORI de resolución comparable al paisaje nucleosomal generado, implementamos un estudio cuantitativo de abundancia de intermediarios de replicación mediante parejas de oligonucleótidos solapantes sobre preparaciones de SNS de tamaño creciente provenientes de tres fracciones consecutivas de gradientes de sacarosa (ver Materiales y Métodos) (**Figura 11**). Representamos los valores de amplificación de cada amplicón normalizados respecto a los obtenidos para una de las regiones flanqueantes situadas a aproximadamente 700-1.000 bp de la región estudiada. Tal y como se esperaría de un patrón de elongación bidireccional a partir del sitio de inicio y con el nivel basal situado en una posición cercana, los enriquecimientos relativos en cada región disminuyen de manera inversamente proporcional a la longitud de los intermediarios de replicación (**Figura 11**).

En los dos ORI más eficientes del estudio, *Mecp2* y *Atp2b3*, localizamos el punto de máximo enriquecimiento en SNS en la posición protegida por el nucleosoma +1 (**Figura 11**). Con el aumento del tamaño de las SNS -aunque disminuía el enriquecimiento relativo- se mantenía este perfil sugiriendo que, en ambos casos, el inicio de la replicación ocurría en el mismo lugar en la mayoría de las células de la población (lo que podría en parte explicar la mayor eficiencia de estos ORI inferida por qPCR). En el ORI asociado al promotor del gen *Hs6st2* el máximo enriquecimiento también coincidía con la región protegida por el nucleosoma +1 en la preparación de SNS más pequeña que evolucionaba a un patrón bimodal en la preparación de 300-800 nt -similar al detectado en la **Figura 7**- con un sutil máximo relativo bajo el nucleosoma -1. Sendos máximos eran inapreciables en la muestra de intermediarios de replicación de mayor tamaño. Estos resultados confirman los obtenidos a escala más global en los perfiles agregados: la síntesis del DNA en los promotores asociados a CGI ocurre en regiones de alta ocupación nucleosomal.

Por último, en los ORI no asociados a promotores *Pias3* y *Syde1*, el perfil de enriquecimiento en intermediarios de replicación se mostraba tan deslocalizado como el entorno nucleosomal (**Figura 11**). Esta correspondencia entre el mayor o menor posicionamiento nucleosomal con el perfil más o menos definido de iniciación de la replicación lo hemos corroborado en otros ORI estudiados en el laboratorio, lo que sugiere la existencia de una relación íntima entre la estructura de la cromatina y los sitios genómicos en los que comienza la síntesis del DNA (Lombraña *et al.*, 2013).

A



B

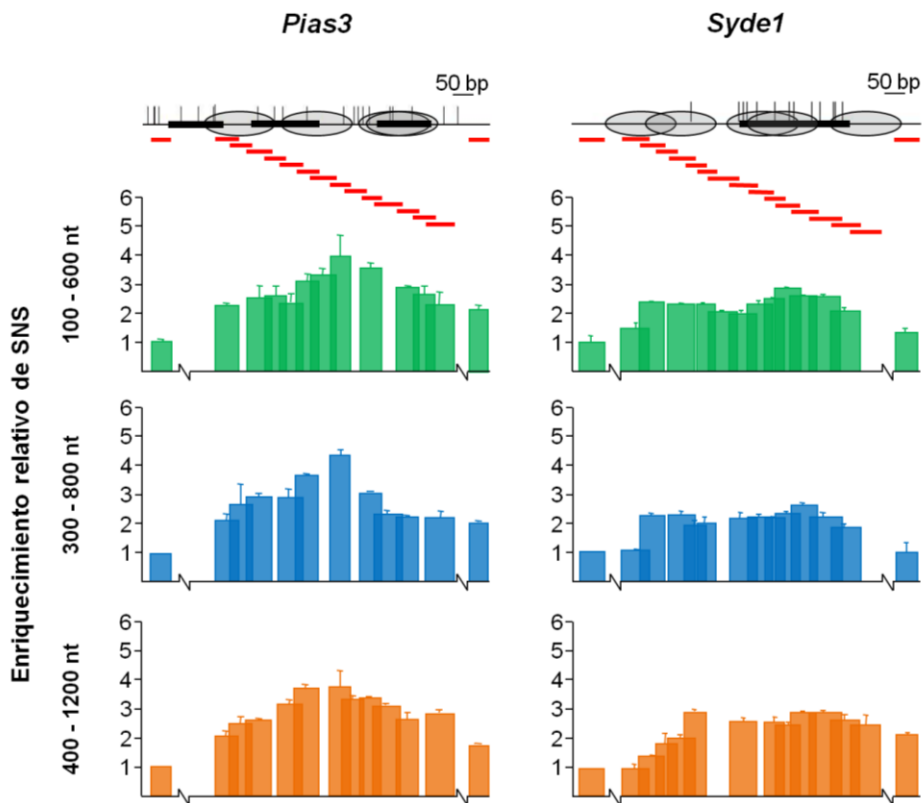


Figura 11: Mapeo de alta resolución del inicio de la replicación en ORI asociados a promotores de CGI (A) o a exones de genes (B). Los histogramas representan los resultados del enriquecimiento de cada uno de los amplicones indicados respecto al flanco más alejado en intermediarios de replicación de tamaño creciente (150-600 nt, 300-800 nt y 400-1200 nt en verde, azul y naranja, respectivamente). La localización de los nucleosomas deriva de las Figuras 9 y 10. En rojo, los amplicones utilizados en el análisis de qPCR.

3. Estudio genómico de la estructura de la cromatina y el inicio de la replicación en células humanas

Los análisis genómicos del inicio de la replicación y de la estructura de la cromatina realizados en células de ratón presentaban algunas limitaciones como el reducido tamaño de la muestra o que la obtención de material biológico para el análisis de la cromatina se hubiera llevado a cabo en condiciones que no mantuviesen la integridad de los nucleosomas más lábiles (Figura 12). Por otra parte, los análisis locales, aunque realizados a una resolución sin precedentes en la literatura, se limitaban a cinco ejemplos -ocho en el estudio completo (Lombraña *et al.*, 2013)-. Por tanto, decidimos realizar un nuevo análisis genómico utilizando los datos de unión de la proteína ORC1 al DNA (Dellino *et al.*, 2013), los datos de localización de variantes de histonas que confieren al nucleosoma en el que se incorporan una alta inestabilidad (Jin *et al.*, 2009) y los datos del mapeo de SNS (Besnard *et al.*, 2012) en células humanas HeLa.

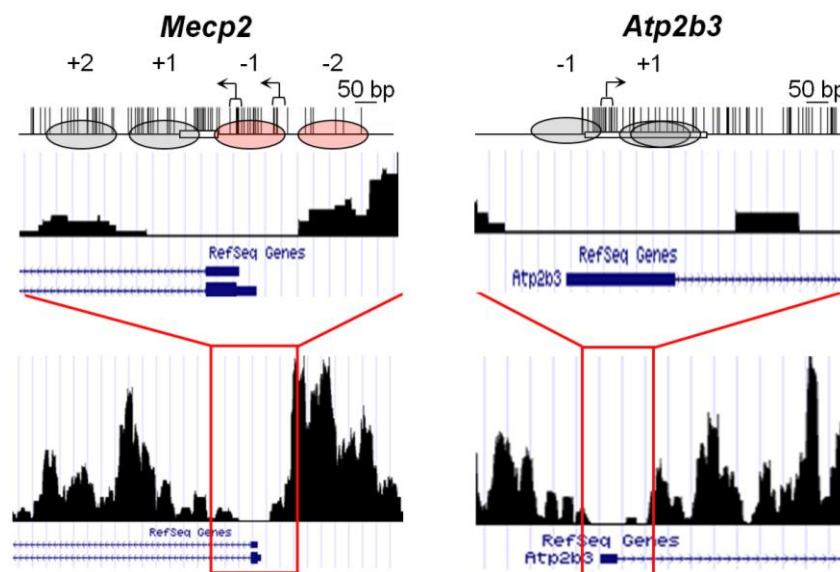


Figura 12: Comparación de los mapas de ocupación nucleosomal obtenidos en las Figuras 9 y 10 mediante NuSA y los derivados de los análisis genómicos a partir de los datos de Teif *et al.* (2012).

En la **Figura 13A** mostramos los resultados de localización de las partículas nucleosomales compuestas por las variantes de histonas H3.3/H2A.Z, que habían sido detectadas únicamente cuando la cromatina se extraía en baja concentración de sal (Jin y Felsenfeld, 2007; Jin *et al.*, 2009), junto a la localización de los nucleosomas compuestos por H2A.Z aislados en alta sal y el posicionamiento de ORC1 en 3.310 CGI. El perfil agregado refleja que la proteína ORC1 se localiza preferentemente 5´ respecto al TSS, exactamente en la posición ocupada por los nucleosomas formados por las variantes de histonas H3.3/H2A.Z, que delimitan la región NDR definida en preparaciones convencionales de mapeo de nucleosomas (Ozsolak *et al.*, 2007; Schones *et al.*, 2008). Estos resultados indican que las regiones de promotores-ORI se encuentran enriquecidas en nucleosomas inestables -como detectamos en la región examinada del gen *Mecp2* y en otros ejemplos del estudio completo (Lombraña *et al.*, 2013)-. El hecho de que la región protegida por nucleosomas inestables coincidiese con la región de unión preferente de ORC1 sugería o bien una posible interacción del complejo ORC con nucleosomas lábiles compuestos por H3.3/H2A.Z, o más probablemente, la existencia de un recambio de los nucleosomas para permitir el acceso del complejo al DNA.

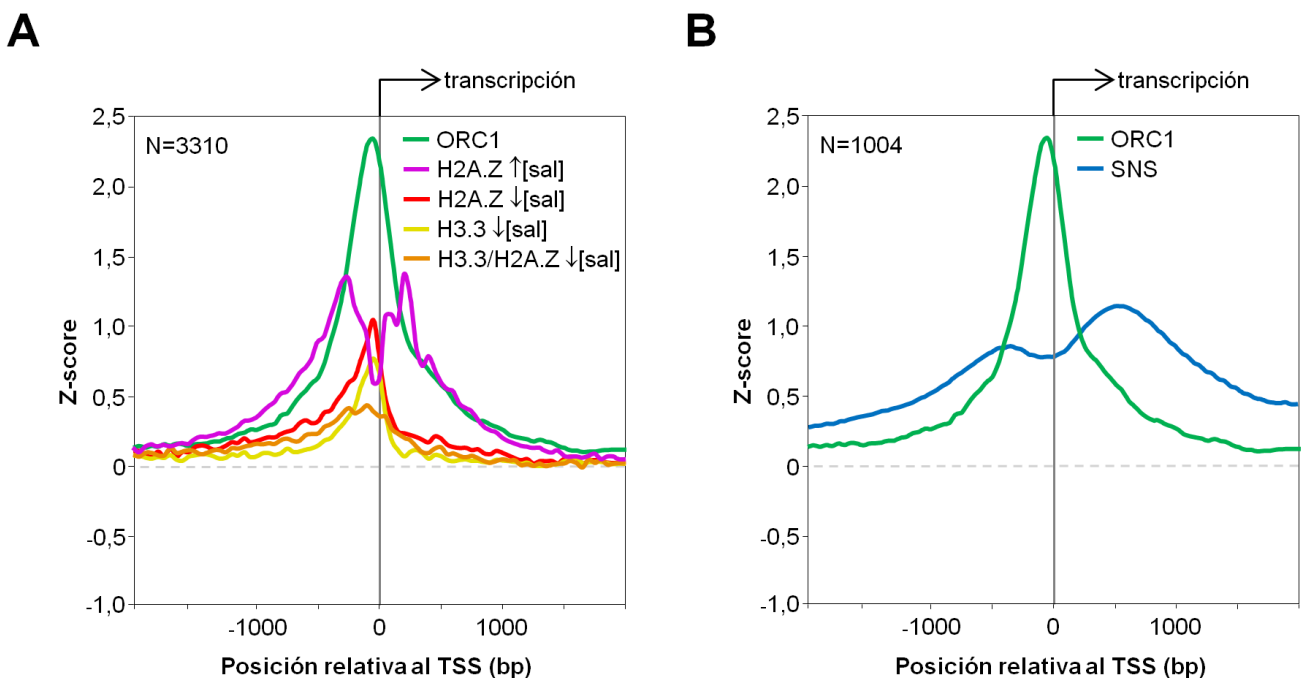


Figura 13: Distribución genómica de los nucleosomas compuestos por las variantes de histonas H2A.Z/H3.3 y los sitios de unión de ORC1 en ORI eficientes en células HeLa. **(A)** Los perfiles agregados representan el Z-score de la densidad de nucleosomas formados por las variantes lábiles de histonas H2A.Z/H3.3 aisladas en baja o alta concentración de sal según se indica (Jin *et al.*, 2009), y de ORC1 (Dellino *et al.*, 2013). Las lecturas se normalizaron en ventanas no solapantes de 50 bp centradas en 3310 TSS positivos para ORC1. **(B)** El mismo análisis se realizó en 1004 TSS positivos para la unión de ORC1 y enriquecidos en SNS (Besnard *et al.*, 2012).

A continuación examinamos la distribución de ORC1 y SNS en el mismo tipo celular (**Figura 13B**). A pesar de que el solapamiento entre ambos grupos de datos era solo del 30%, las CGI positivas para la unión de ORC1 y clasificadas como ORI (Besnard *et al.*, 2012), mostraron una localización claramente distinta de ORC1 y de los sitios de iniciación de la replicación. De modo similar al patrón observado en células de ratón (**Figura 7B**), el perfil bimodal de replicación detectado en células HeLa se distribuye a ambos lados del TSS, coincidiendo con las regiones de mayor ocupación nucleosomal.

En trabajos recientes de mapeo genómico de ORI en células de mamíferos se ha puesto de manifiesto que un alto porcentaje de los sitios de iniciación identificados, tanto en células de ratón como humanas, contienen motivos de secuencias con el potencial de formar estructuras de G4 (Cayrou *et al.*, 2012; Besnard *et al.*, 2012; Comoglio *et al.*, 2015). El papel de estas estructuras en la especificación de los ORI no está clara, pero evidencias recientes demuestran que algunos de estos motivos pueden estimular la iniciación de la replicación (Valton *et al.*, 2014). Una posibilidad interesante es que las estructuras de G4 pudieran facilitar el reclutamiento de ORC a los ORI, ya que se ha demostrado en ensayos *in vitro* que el complejo ORC humano tiene afinidad por los motivos G4 a través de un dominio de la proteína ORC1 (Hoshina *et al.*, 2013). Esta afinidad, además, está muy incrementada cuando los motivos G4 están presentes en DNA de cadena sencilla o en RNA. Decidimos, por tanto, estudiar la asociación entre los sitios de unión de ORC1 y la abundancia de motivos G4 en promotores-ORI -muy frecuentes en CGI- en los que se había detectado la presencia de estructuras persistentes de DNA de cadena sencilla concomitantes con la transcripción o *R-loops*. Estas estructuras de tres cadenas de ácidos nucleicos son termodinámicamente muy estables y se generan por la hibridación de una molécula de RNA a la cadena de DNA complementaria dejando desplazada la segunda hebra de DNA en forma de cadena sencilla (Roberts y Crothers, 1992). Su presencia *in vivo* en el genoma humano se correlaciona con regiones asimétricas en la distribución de G/C (*GC-skew*) -la cadena no-molde para la transcripción tiene un exceso de G sobre C-, lo que ocurre principalmente en las posiciones 5' y 3' de los genes y, específicamente, en el 75% de todos los promotores asociados a CGI (Ginno *et al.*, 2012; 2013).

Para analizar la asociación entre estos tres parámetros, seleccionamos aquellos TSS situados en CGI que, siendo positivos para ORC1 y clasificados como ORI por su enriquecimiento en SNS -33% del total definido por Dellino *et al.* (2013)-, generaban *R-loops in vivo* -evaluado por experimentos de DRIP-seq

(inmunoprecipitación con anticuerpos específicos frente a híbridos DNA:RNA acoplados a secuenciación masiva) (Ginno *et al.*, 2013)-. En primer lugar, definimos un subgrupo de 1.953 CGI con ORC1 asociado a ORI (CGI con ORC1/ORI) que, aun siendo muy restrictivo (20% del total de 9.864 CGI), debería conformar un conjunto de ORI de alta eficiencia por las características de los mismos (Sequeira-Mendes *et al.*, 2009; Besnard *et al.*, 2012). Por otro lado, seleccionamos las 1.571 CGI coincidentes con regiones positivas para DRIP-seq (CGI con DRIP) -lo que suponía el 40% de todas las regiones de *R-loops* identificadas en Ginno *et al.* (2013)-. El solapamiento entre estos dos subgrupos era significativo (**Figura 14A**): el 30% de las CGI con presencia de *R-loops* exhibían unión de ORC1 inmediatamente al 5' de su TSS (la asociación esperada al azar sería del 6%) (**Figura 14B**). A su vez, la señal de DRIP y de enriquecimiento en ORC1 solapaban precisamente en la región de 1 Kb con GC-skew positivo identificado en este grupo de islas (Ginno *et al.*, 2012) (**Figura 14B**).

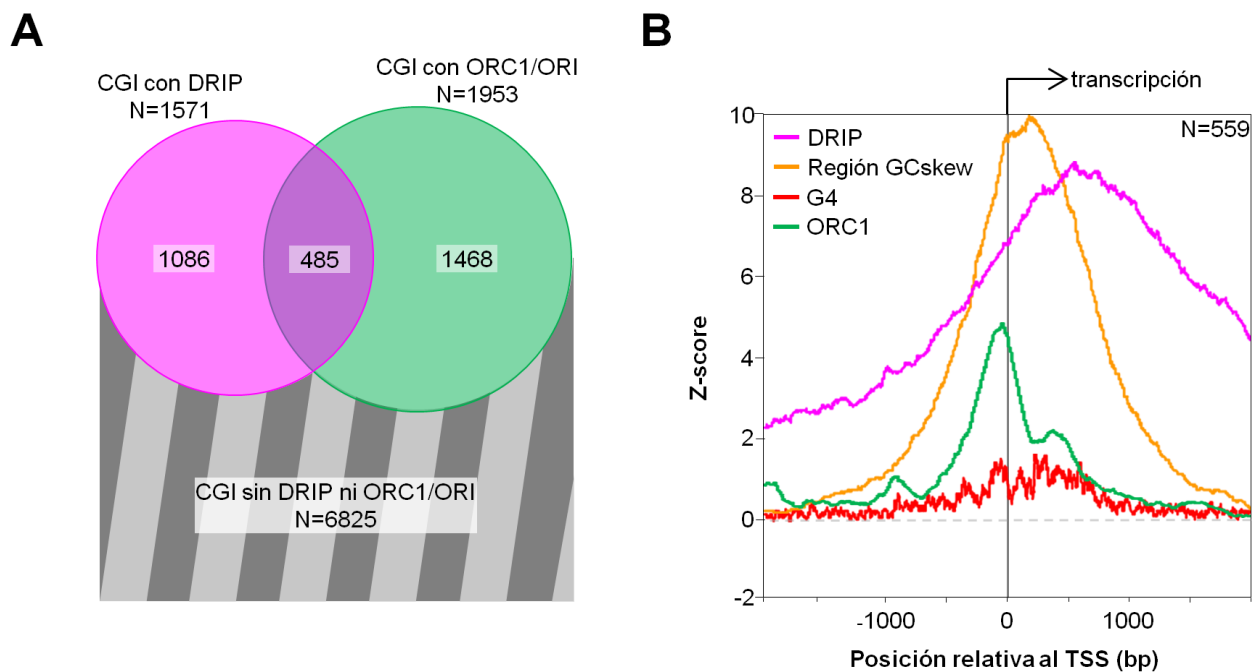


Figura 14: Relación entre *R-loops* y sitios de unión de ORC1 en ORI asociados a CGI en células humanas. **(A)** Diagramas de Venn ilustrativos de la asociación entre *R-loops* -definidos como regiones consenso en réplicas de DRIP-seq en Ginno *et al.* (2013)- y sitios de unión de ORC1 en ORI -definidos cruzando los datos de Dellino *et al.* (2013) y Besnard *et al.* (2012) en CGI-. **(B)** Perfiles agregados de los Z-score correspondientes a la distribución de regiones con *R-loops*, regiones con GC-skew (Ginno *et al.*, 2012), sitios de unión de ORC1 y motivos G4 en el grupo de promotores-ORI definidos en (A) centrados en los TSS de los 485 ORI asociados a CGI con presencia de *R-loops* y ORC1. La posición de los motivos G4 se determinó aplicando el software Quadparser especificando un tamaño del bucle de 1-7 nucleótidos (Huppert y Balasubramanian, 2005).

Aun teniendo en cuenta que solo con los resultados de DRIP-seq no podía conocerse la localización exacta de los *R-loops* en cada promotor -ya que su resolución depende de la distribución de los sitios de restricción utilizados para fragmentar el genoma-, que los experimentos de DRIP y del mapeo de ORI habían sido realizados en tipos celulares distintos (células NTera2 y HeLa, respectivamente) y el elevado grado de restricción utilizado para elegir los grupos de datos, estos análisis muestran que los *R-loops* co-localizan con los sitios de unión de ORC1 en una fracción de los ORI más eficientes en células humanas. Y, lo que es más importante, el enriquecimiento en motivos G4 en estas regiones es paralelo a la señal de unión de ORC1 (**Figura 14B**), asociación consistente con la idea de que las estructuras de G4 podrían contribuir al reclutamiento de ORC a esas regiones específicas.

PARTE II

1. Anotación del genoma de *L. major*

En las secciones anteriores hemos estudiado la relación del inicio de la replicación con la organización de la cromatina y con estructuras no canónicas del DNA en una categoría de ORI de mamíferos, los asociados a regiones promotoras de CGI. Para evaluar nuestra hipótesis de que la distribución genómica de los ORI está integrada con la actividad transcripcional en general, elegimos un sistema más sencillo con una particular organización genómica: el parásito humano *L. major*. Los protozoos parásitos como *Leishmania* y *Trypanosoma* se separaron pronto en la línea evolutiva de los eucariotas (revisado en Baldauf, 2003). En ellos se han realizado numerosos descubrimientos fundamentales para entender la evolución de procesos moleculares básicos y, especialmente, la simplificación y especialización de los componentes de la maquinaria de transcripción para la vida parásita (Ivens *et al.*, 2005).

El genoma de *L. major* tiene un tamaño de 32,8 Mb repartidas en 36 cromosomas (de un tamaño comprendido entre 0,28 Mb y 2,8 Mb). Estos cromosomas, relativamente pequeños, no parecen tener regiones subtelméricas ni centrómeros. Una de las peculiaridades de estos eucariotas unicelulares es que los genes codificantes para proteínas se agrupan en largas unidades policistrónicas de transcripción o PTU que son transcritas por la RNAPII a lo largo de la misma cadena de DNA y que pueden contener decenas o centenares de genes (**Figura 15**) (Martinez-Calvillo *et al.*, 2004; Ivens *et al.*, 2005). Las regiones de cambio de cadena -*Strand Switch Region* (SSR)-, en las que la polaridad transcripcional cambia, pueden ser divergentes (SSRd) o convergentes (SSRc) para la transcripción. En las primeras, la transcripción comienza en direcciones opuestas y en las segundas, la transcripción termina y se desacopla la RNAPII del DNA (Martinez-Calvillo *et al.*, 2003; 2004). Existe un tercer tipo de regiones en las que la transcripción termina y comienza en la misma cadena sin cambiar la polaridad (van Luenen *et al.*, 2012) y que designamos como *Head-Tail* (HT).

Los largos pre-RNAm de las PTU son procesados co-transcripcionalmente mediante la poliadenilación del extremo 3' y el ensamblaje por *trans-splicing* de una secuencia líder (SL) en el 5' del siguiente gen, generando los RNAm maduros (**Figura 15**) (Agabian, 1990; Agami y Shapira, 1992; LeBowitz *et al.*, 1993). Por tanto, la regulación de la expresión génica en *L. major* ocurre principalmente de modo post-

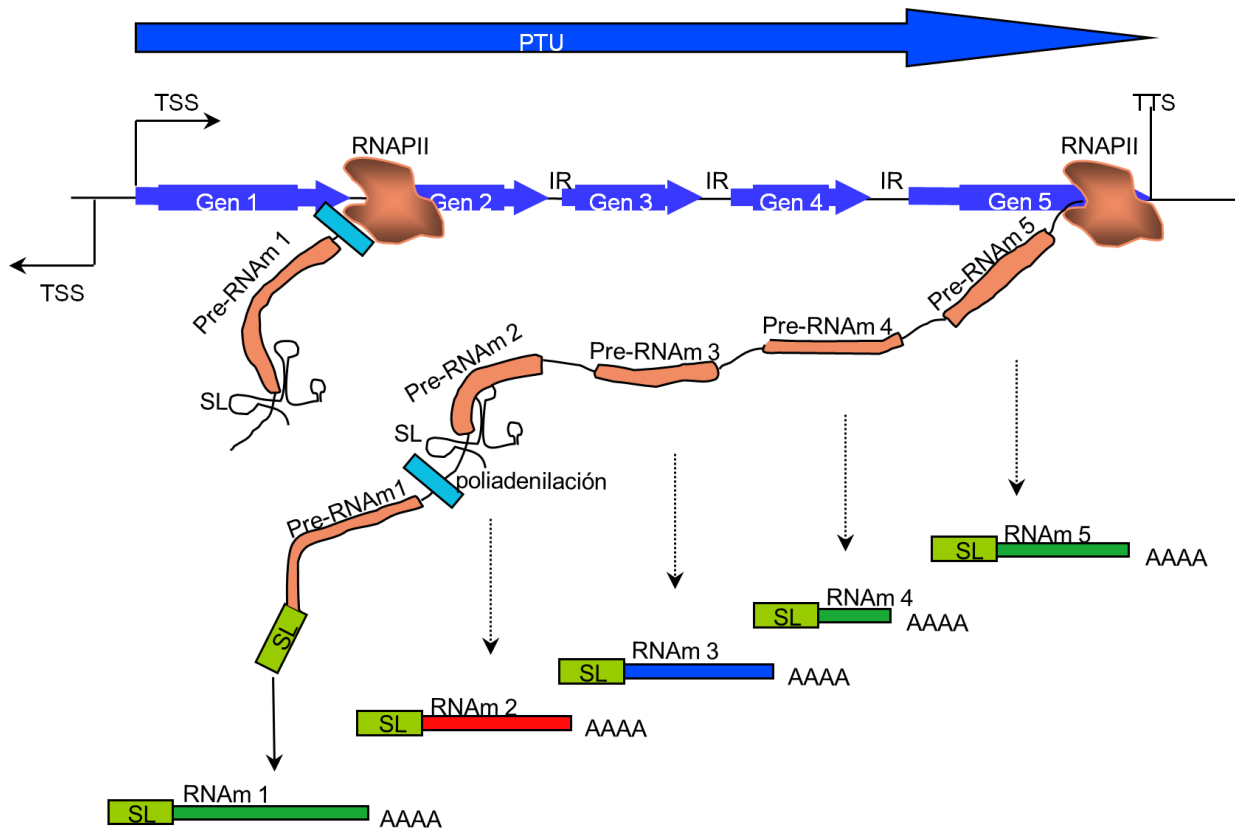
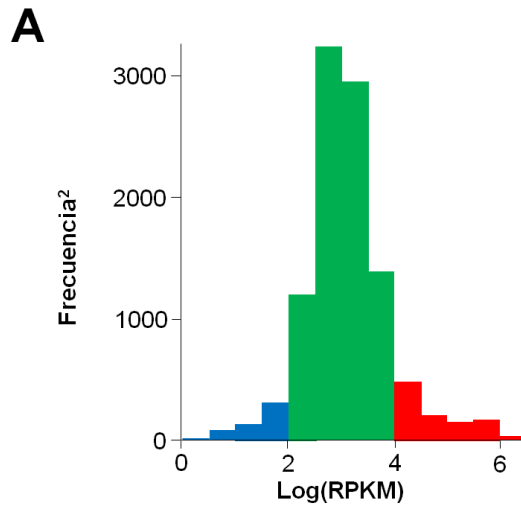


Figura 15: Esquema de la organización génica de *L. major* así como de su sistema transcripcional. Los genes se agrupan en unidades policistrónicas de transcripción (PTU) que son transcritas la RNA Polimerasa II (RNAPII). De manera co-transcripcional se produce el *trans-splicing* que lleva a la maduración de los pre-RNAm por medio del acoplamiento de una secuencia líder (SL) y poliadenilación. La estabilidad de los RNAm maduros modula la expresión génica del parásito puesto que la mayoría de la regulación parece ocurrir a nivel post-transcripcional.

transcripcional, mediante la estabilización diferencial de los RNAm maduros o a nivel pre- y post-traducciona (revisado en Stiles *et al.*, 1999; Requena, 2011). Otra singularidad de estos eucariotas es la presencia de una base modificada en sustitución de algunas timinas: el β -D-glucosil-hidroximetiluracilo o base J. Esta base se sintetiza a través de una oxidación de timina seguida de glucosilación (Gommers-Ampt *et al.*, 1993). La base J sustituye en torno al 1% de la timina del genoma, localizándose principalmente en los telómeros (aproximadamente el 99%), mientras que el 1% restante se reparte entre las SSRc y en la proximidad de los *loci* ribosomales y de las secuencias SL. La pérdida de la base J produce defectos en la terminación de la transcripción en las regiones SSRc y HT y genera RNA antisentido, lo que sugiere que esta base modificada podría tener un papel en la terminación de la transcripción (van Luenen *et al.*, 2012).



B

Regiones definidas	Número de casos	Extensión total (bp)	Proporción genómica (%)
Cambio de polaridad convergente (SSRc)	37	85.240	0,27 %
Cambio de polaridad divergente (SSRd)	55	121.930	0,39 %
Fin-comienzo de transcripción (HT)	51	116.724	0,37 %
3'-UTR	7.517	6.959.685	22,34 %
5'-UTR	7.558	3.411.339	10,95 %
Inter pre-RNAm (IR)	7.134	2.754.914	8,84 %
Regiones codificantes (CDS)	7.508	14.244.783	45,72 %
Zonas de sombra	169	1.695.340	5,16 %
Otras discrepancias de anotación	-	1.769.800	5,68 %

Figura 16: **A:** Distribución de los RNAm poliadenilados en función de su estabilidad [log(RPKM)] y clasificación en base a ésta en: baja (azul), media (verde) o alta (roja). **B:** Regiones anotadas del estudio en *L. major* indicándose el número de ellas así como su extensión en bp y la proporción del genoma que ocupan

Como resultado de esta peculiar organización génica la mayoría del genoma de *L. major* se transcribe constitutivamente por la RNAPII y los promotores y sitios de terminación (TTS) constituyen una pequeña fracción del genoma que además, se encuentra muy localizada. De este modo, este sistema representaba un escenario ideal para estudiar la relación entre los procesos de replicación y transcripción. Para realizar este estudio era necesario integrar todos los datos genómicos presentes en la literatura: acetilación de histona H3 (AcH3) (Thomas *et al.*, 2009), base J (van Luenen *et al.*, 2012), anotación y estabilidad de RNAm poliadenilados (Rastrojo *et al.*, 2013) y CDS (GeneDB; Logan-Klumpler *et al.*, 2012) junto con nuestros propios datos experimentales de secuenciación masiva de SNS (SNS-seq) o de fragmentos de DNA resistentes a la digestión por la enzima MNasa (MNasa-seq). Además, en los estudios del transcriptoma de *L. major*, conjuntamente a la descripción de los pre-RNAm del parásito, se había determinado la abundancia relativa de cada uno de ellos, que por las peculiares características de regulación transcripcional de este protozoo, la mayor o menor persistencia temporal de los RNAm se asocia a su estabilidad molecular (Rastrojo *et al.*, 2013). De este modo, dividimos la abundancia relativa de los RNAm (log RPKM) en tres intervalos: baja (0-2), media (2-4) y alta (mayor de 4) (**Figura 16A**)

Como paso previo al análisis computacional de todos estos datos genómicos, tuvimos que realizar una curación manual del genoma ya que, al alinear las lecturas del DNA genómico control de los experimentos de SNS-seq y MNasa-seq, encontramos muchas regiones con drásticos incrementos o ausencias de lecturas que no podían explicarse únicamente en base a secuencias repetitivas, duplicaciones del genoma no anotadas o artefactos del alineamiento. De este modo, definimos 169 regiones (aproximadamente, el 5% del genoma) como “zonas de sombra” que descartamos de los análisis posteriores (**Anexo I: Zonas excluidas del análisis**). También detectamos algunos conflictos entre las anotaciones de algunos RNAm y sus correspondientes CDS como la presencia de CDS sin RNAm asociado o viceversa (lo que nos impedía definir con precisión las regiones UTR de los genes). En la **Figura 16B** se resumen los resultados de esta curación del genoma de *L. major*.

Además de por la direccionalidad de transcripción de las PTU, las 55 SSRd previamente identificadas eran fácilmente reconocibles por la presencia de una prominente señal de AcH3 que marcaba los TSS de la RNAPII (Thomas *et al.*, 2009) (**Figura 17**, posición 735 Kb). Por otra parte, en 36 de las 37 SSRc anotadas detectamos la presencia de base J (van Luenen *et al.*, 2012) (**Figura 17**, posición 640 Kb) que marca los TTS de las PTU. Para identificar las regiones HT, que no estaban anotadas en *L. major*, cruzamos los datos del transcriptoma (Rastrojo *et al.*, 2013) con

la localización de señales de base J y de AcH3 -descartando las correspondientes a SSR-. De este modo definimos 51 regiones HT, caracterizadas por la ausencia de actividad transcripcional y que presentaban señal de enriquecimiento en base J seguido de señal de enriquecimiento en AcH3 en la misma hebra (**Figura 17**, posición 550 Kb).

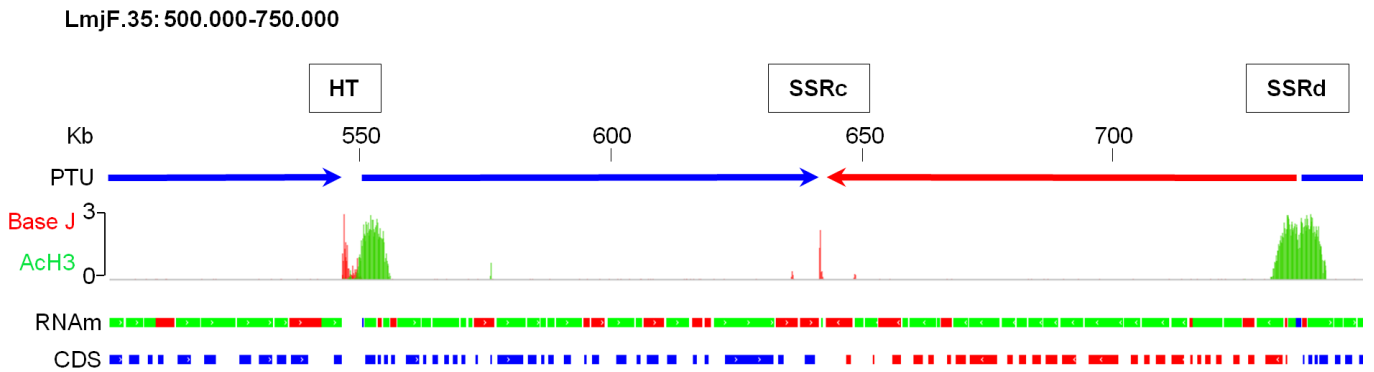


Figura 17: Ejemplos de los tres tipos de regiones inter-PTU del genoma de *L. major*: cambio de hebra divergente con inicio de transcripción (SSRd) en la posición 735 Kb, cambio de hebra convergente con final de transcripción (SSRc) en la posición 640 Kb y, terminación y comienzo de la transcripción en la misma hebra (HT) en la posición 550 Kb. Las PTU se representan en azul o rojo dependiendo de la hebra de DNA que se utilice como molde (hebra + o hebra – respectivamente). En la siguiente línea se representan los datos de localización de la base J en rojo (van Luenen *et al.*, 2012) y de acetilación de H3 en verde (Thomas *et al.*, 2009) que marcan los TTS y TSS de las PTU, respectivamente. En las dos últimas líneas representamos los RNAm (Rastrojo *et al.*, 2013) en rojo, verde o azul en función de su abundancia (alta, media o baja respectivamente) y las regiones codificantes en azul o rojo según el sentido de la hebra codificante.

Dada la asociación entre los ORI y los motivos G4 encontrada en estudios genómicos de varios sistemas animales como células humanas (Besnard *et al.*, 2012), de ratón (Cayrou *et al.*, 2012) y de *Drosophila* (Comoglio *et al.*, 2015), resultaba de interés analizar la distribución de estos motivos en el genoma de *L. major*, que contiene un 63% de G/C. Para ello utilizamos el programa *Quadparser* (Huppert y Balasubramanian, 2005) que identificó más de 11.000 candidatos que respondían a la estructura general $G_3-L_{1-7}-G_3-L_{1-7}-G_3-L_{1-7}-G_3$. La mayoría de los motivos G4 se localizaban en las regiones 5'-UTR y 3'-UTR de los genes, así como en las regiones intergénicas (IR) (**Figura 18A**). Sin embargo, cuando analizamos la densidad de motivos G4 teniendo en cuenta el tamaño de las distintas regiones en el genoma, encontramos que las regiones de terminación de la transcripción (tanto SSRc como HT) mostraban hasta el doble de densidad de estos motivos que las demás (**Figura 18B**).

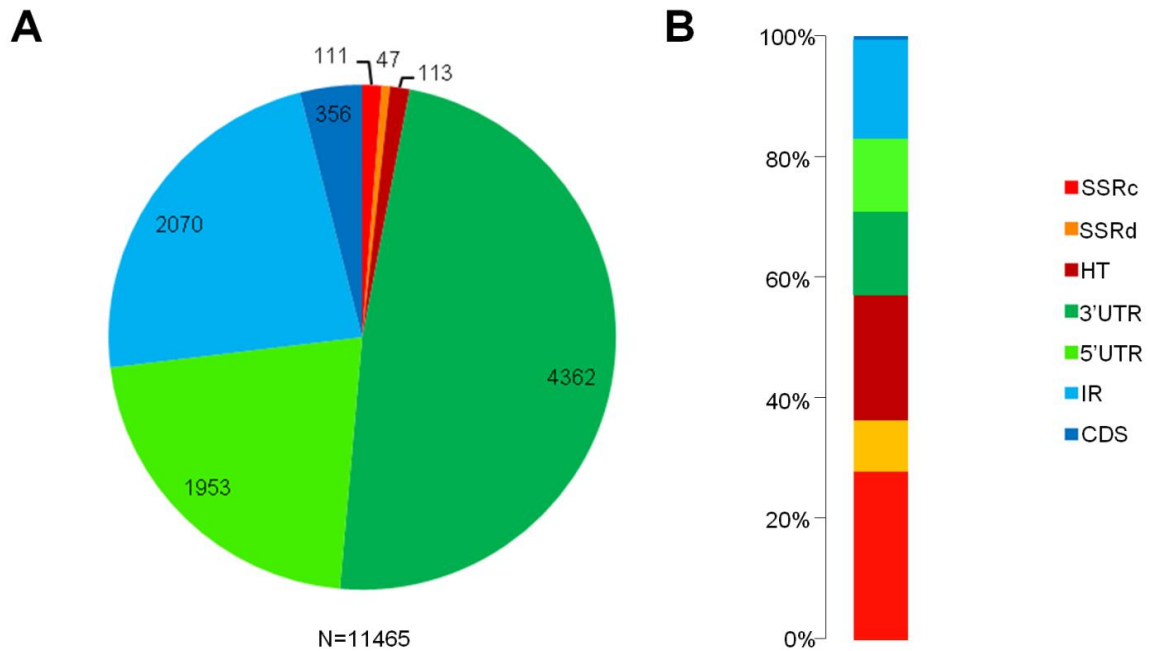


Figura 18: Distribución genómica de motivos G4 en *L. major* a lo largo de las diferentes regiones definidas para este trabajo, tanto en número absoluto (A), como en densidad en función del tamaño de cada región (B). La densidad de estos motivos es muy alta en zonas de terminación de la transcripción (SSRc y HT). También resulta destacable su presencia en las zonas no codificantes de las PTU (UTR e IR).

2. Análisis genómico de la iniciación de la replicación en *L. major*

Tal y como se ha descrito para los análisis realizados en células de mamíferos, para el mapeo genómico de los sitios de iniciación de la replicación de *L. major* purificamos intermediarios de replicación de fracciones consecutivas de gradientes de sacarosa, preparados a partir de dos muestras biológicas independientes de cultivos de células promastigotes en fase exponencial de crecimiento y, posteriormente, procedimos a su secuenciación masiva (ver Materiales y Métodos). En una primera serie de experimentos utilizamos la fracción que contenía intermediarios de replicación entre 300-800 nt. Como comentamos en la sección de Materiales y Métodos, existe cierta controversia en el campo sobre el uso de la enzima λ -exonucleasa en el protocolo de enriquecimiento de intermediarios de replicación con un cebador de RNA (Fouk *et al.*, 2015) a pesar de ser la metodología más extendida (Cadoret *et al.*, 2008; Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011; Besnard *et al.*, 2012; Picard *et al.*, 2014; Cayrou *et al.*, 2015; Comoglio *et al.*, 2015). Esta crítica se basa en la hipotética

dificultad del enzima para digerir secuencias ricas en C/G o estructuras no canónicas del DNA como los G4. Para minimizar el sesgo putativo de digestión preferente de secuencias ricas en A/T respecto a secuencias ricas en G/C, introdujimos una variación al protocolo estándar que consistió en realizar tres rondas sucesivas de fosforilación con polinucleótido quinasa (PNK) y digestión con λ -exonucleasa (Comoglio *et al.*, 2015). Además, comparamos los resultados obtenidos utilizando todas las lecturas de secuenciación con los obtenidos tras eliminar computacionalmente aquellas lecturas cuyo alineamiento coincidiese con secuencias con posibilidad de formar estructuras de G4.

A

Programa de alineamiento	Todas las lecturas	Lecturas sin G4
Bowtie	815.557	708.337
BWA (en RUBioSeq)	857.159	603.949

B

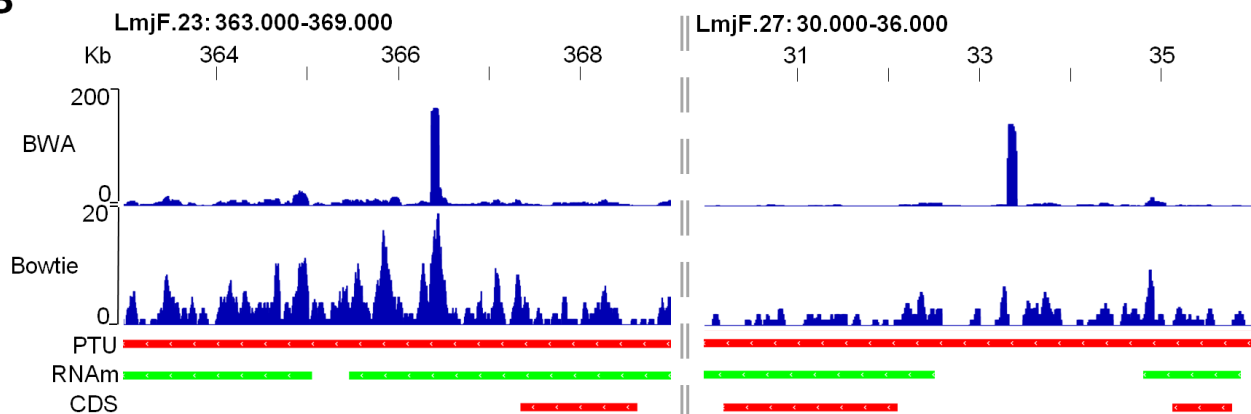


Figura 19: Comparación de los resultados obtenidos utilizando dos alineadores diferentes para las lecturas de la primera réplica de SNS-seq en *L. major*. **A:** Profundidad de lecturas alineadas con los dos programas de utilizados tanto antes como después del filtrado de aquellas lecturas que contenían motivos G4 en su extremo 5'. **B:** Ejemplos representativos de dos regiones del genoma de *L. major* en las que detectamos artefactos de alineamiento con BWA (perfil superior) respecto a Bowtie (perfil inferior). Nótese la diferencia de hasta un orden de magnitud en los artefactos provocados por el primer alineador. En las líneas inferiores representamos las unidades de transcripción (PTU) en azul (hebra +) o rojo (hebra -), los RNAm (Rastrojo *et al.*, 2013) en rojo, verde o azul en función de su abundancia (alta, media o baja) y los CDS en azul o rojo según el sentido de la hebra codificante.

Para el alineamiento de las lecturas probamos dos alineadores diferentes: Bowtie y BWA. El primero es ampliamente utilizado en trabajos de todo tipo que impliquen secuenciación masiva (Quintales *et al.*, 2014; Giannoulatou *et al.*, 2014; Comoglio *et al.*, 2015). El segundo, forma parte del paquete RUBioSeq, un protocolo de herramientas de software mucho más amplio (Rubio-Camarillo *et al.*, 2013) y, aunque es menos versátil en su manejo, puede alcanzar mejores cotas de efectividad en determinados casos. El número de lecturas que pudimos alinear con BWA fue ligeramente superior a las alineadas con Bowtie (5,1%, de mejora). Sin embargo, el porcentaje de alineamiento decayó un 14,7% cuando descartamos las lecturas que contenían motivos G4 en sus extremos (**Figura 19A**). Tras un análisis minucioso de los perfiles resultantes, también pudimos observar diversos artefactos por acumulación de lecturas repetidas en algunas regiones utilizando BWA (**Figura 19B**). Basándonos en estas observaciones, decidimos utilizar el software Bowtie para el alineamiento de las lecturas provenientes de intermediarios de replicación en *L. major*.

El siguiente paso era determinar cuáles de los lugares con enriquecimiento en lecturas provenientes de intermediarios de replicación eran significativos. Es el proceso conocido como *peak calling*. La dificultad con la que nos encontramos era que la mayoría del software disponible estaba dirigido para el análisis de lecturas provenientes de experimentos de ChIP-seq como, por ejemplo, modificaciones de histonas o factores de transcripción (Xu *et al.*, 2010; Zhang *et al.*, 2008 y revisado en Park, 2009; Mundade *et al.*, 2014). La aplicación de estos algoritmos para el mapeo genómico de la iniciación de la replicación se ha debatido en un trabajo reciente (Picard *et al.*, 2014). Por ello decidimos comparar las dos herramientas más habituales para determinar la que mejor respondía a las particularidades del estudio y del modelo: (a) MACS -el más habitual en análisis de ChIP-seq (Zhang *et al.*, 2008)- y (b) SNS-scan -una serie de algoritmos creados exclusivamente para el análisis de SNS-seq (Picard *et al.*, 2014)-.

Como mostramos en la **Figura 20A**, no encontramos diferencias significativas en la distribución de los ORI con ninguno de los dos algoritmos de *peak calling* utilizados cuando empleamos todas las lecturas para el análisis o tras filtrar aquellas que contenían motivos G4 en alguno de sus extremos. Sin embargo, si observamos una disminución importante en el número de ORI detectados, fenómeno que achacamos a la menor profundidad de lecturas resultantes tras el filtrado. Lógicamente, la mayor parte de los ORI que se perdían pertenecía a las regiones que tienen una proporción mayor de motivos G4. Era especialmente llamativa la falta de detección de ORI en regiones SSRd utilizando el software SNS-scan tras el filtrado.

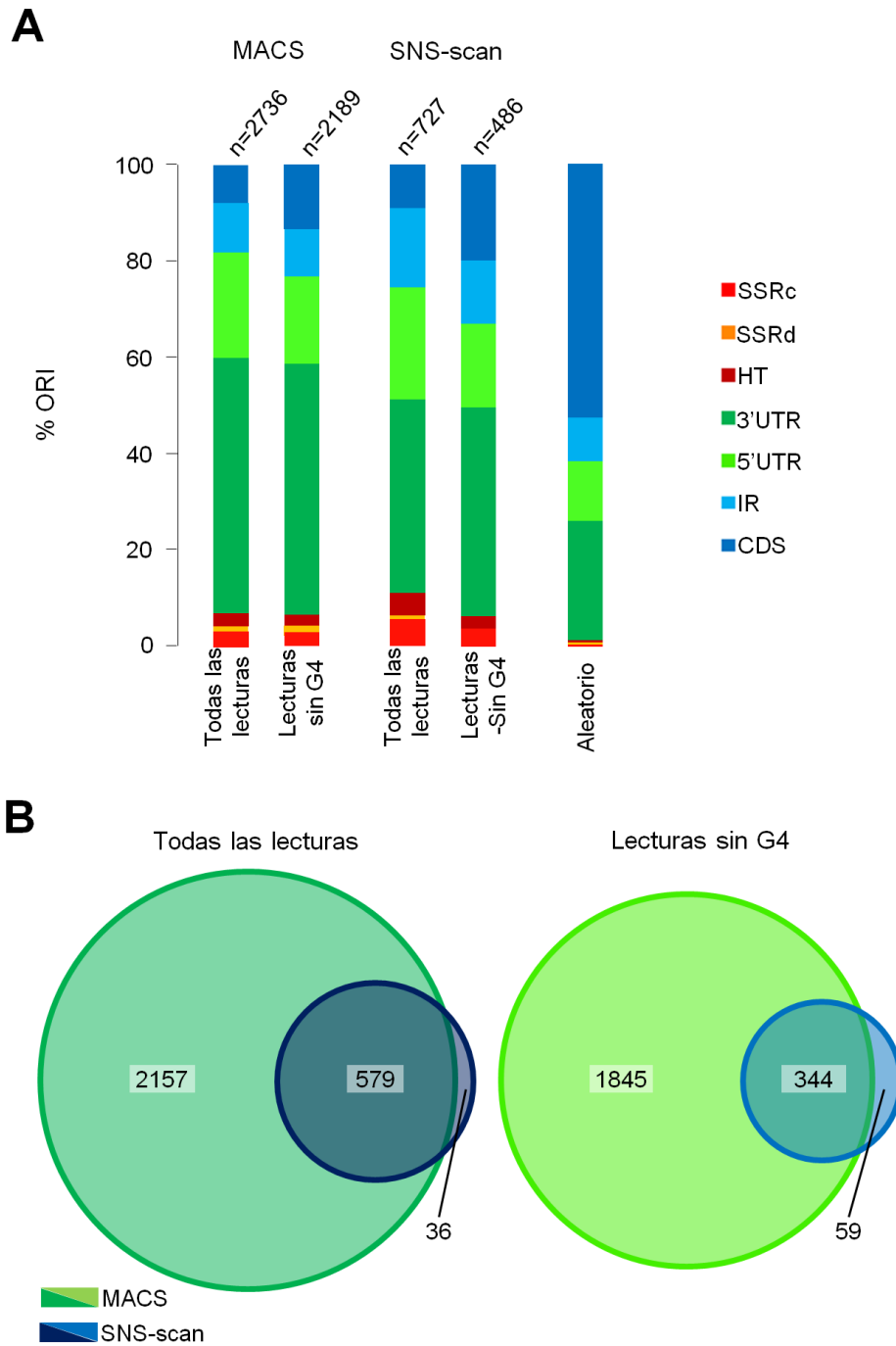


Figura 20: Comparación de los algoritmos MACS y SNS-scan para la identificación de ORI en *L. major* en la primera réplica del experimento de SNS-seq: **(A)** Distribución genómica de los ORI definidos utilizando todas las lecturas alineadas o únicamente aquellas que no contenían motivos G4 en sus extremos en las regiones genómicas definidas en la **Figura 16B** . El histograma de la derecha representa el porcentaje de localización esperada si los ORI se distribuyeran aleatoriamente a lo largo del genoma. **(B)** Análisis del solapamiento entre los ORI definidos por ambos algoritmos.

Por otra parte, el análisis mediante SNS-scan era más restrictivo que MACS a la hora de definir como significativo un enriquecimiento debido, probablemente, a un alto número de falsos negativos. En cualquier caso, era destacable que cualquiera de los análisis arrojaba un resultado cualitativamente similar de distribución de los ORI a lo largo del genoma y, sobre todo, muy diferente al de una distribución aleatoria. Por otra parte, SNS-scan solamente encontraba un número residual de ORI que no hubieran sido detectados por MACS (**Figura 20B**). Teniendo en cuenta estos datos, además de la generalidad de su uso en la literatura y la mayor flexibilidad en la definición de parámetros, decidimos emplear el algoritmo MACS para la identificación de los ORI en *L. major*.

Al analizar las regiones enriquecidas en SNS de este primer experimento de SNS-seq destacaba el hecho de que más del 70% de las zonas de terminación de la transcripción (HT y SSRc) contenían sitios de iniciación de la replicación (**Figura 21A**). Aunque esta proporción decrecía ligeramente al eliminar del análisis las lecturas que solapaban con motivos G4 (histograma azul claro), los enriquecimientos eran altamente significativos cuando se comparaban con los esperados en una distribución aleatoria de ORI ($p < 0,0001$). A continuación analizamos la asociación entre ORI y motivos G4, ya que, como hemos expuesto en puntos anteriores, se ha postulado su

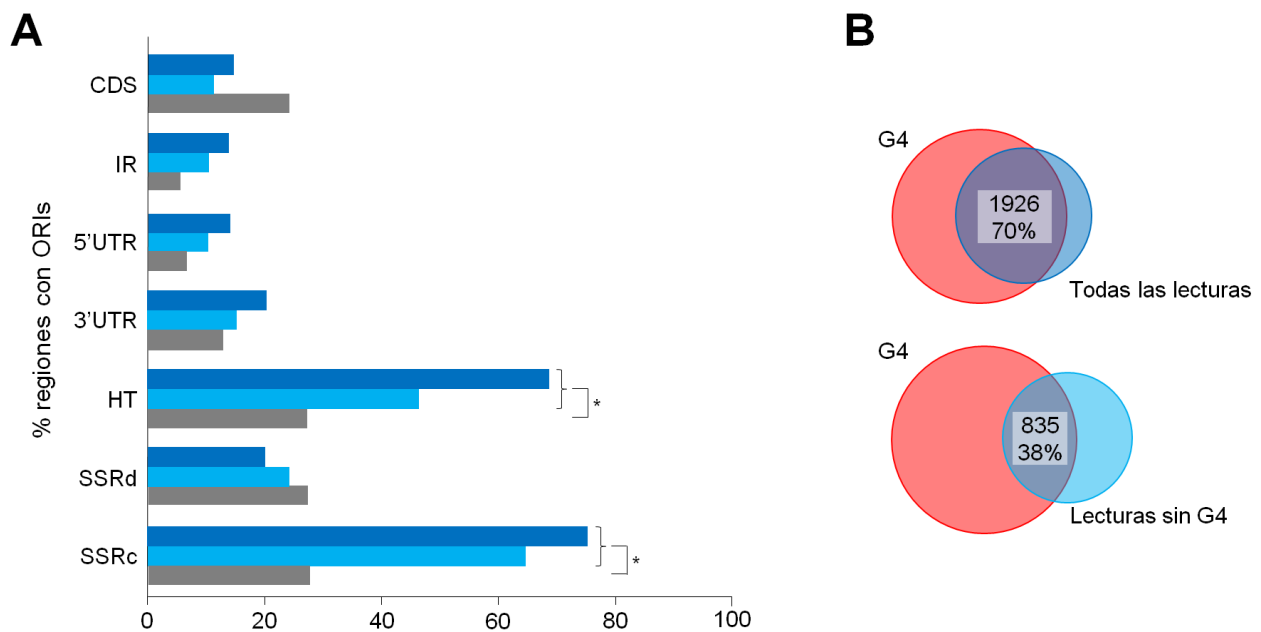


Figura 21: Caracterización de los ORI definidos en la primera réplica del estudio en *L. major*. **(A)** Porcentaje de las regiones que presentaban ORI. En azul oscuro, los ORI definidos con todas las lecturas; en azul claro tras filtrar aquellas lecturas que solapaban con motivos G4 en sus extremos. En gris, distribución esperada al azar ($p < 0,0001$). **(B)**: Solapamiento de los picos de ORI con motivos G4.

co-localización en varios sistemas analizados (Cayrou *et al.*, 2012; Besnard *et al.*, 2012; Valton *et al.*, 2014; Comoglio *et al.*, 2015). En consonancia con estos trabajos, encontramos que un 70% de las regiones de ORI (definidos como segmentos de 800 bp a partir del punto de máximo enriquecimiento en SNS) solapaban con motivos G4 en *L. major* (**Figura 21B**). Aún más, esta asociación se mantenía en un 38% cuando se consideraban sólo los ORI identificados tras eliminar del análisis las lecturas que podrían formar estructuras de G4 en sus extremos.

Dada la baja profundidad de lecturas alineadas en este primer experimento decidimos generar un segundo experimento de SNS-seq purificando intermediarios de replicación de fracciones consecutivas de gradientes de sacarosa: 300-800 nt y de 400-1.200 nt. En este segundo experimento pudimos alinear 2.425.415 y 2.183.228 lecturas respectivamente. De esta forma, y procesando los datos de manera similar al caso anterior, pudimos definir 5.807 y 4.357 ORI en las respectivas fracciones. Era interesante destacar que, a pesar de haber sido tratadas independientemente, el 73% de picos de la fracción de 300-800 nt eran detectados en la fracción que contenía intermediarios de replicación de mayor tamaño, mientras que el 88% de los picos definidos en la fracción mayor también se detectaban en la primera (**Figura 22**). Además, al realizar la comparación con el experimento anterior, el 84% de aquellos ORI eran detectados en las nuevas réplicas. Por tanto, consideramos válida la identificación de los ORI en este modelo mediante dos réplicas independientes, siendo mucho más sensible este segundo grupo de experimentos. Así pues, procedimos a su análisis detallado.

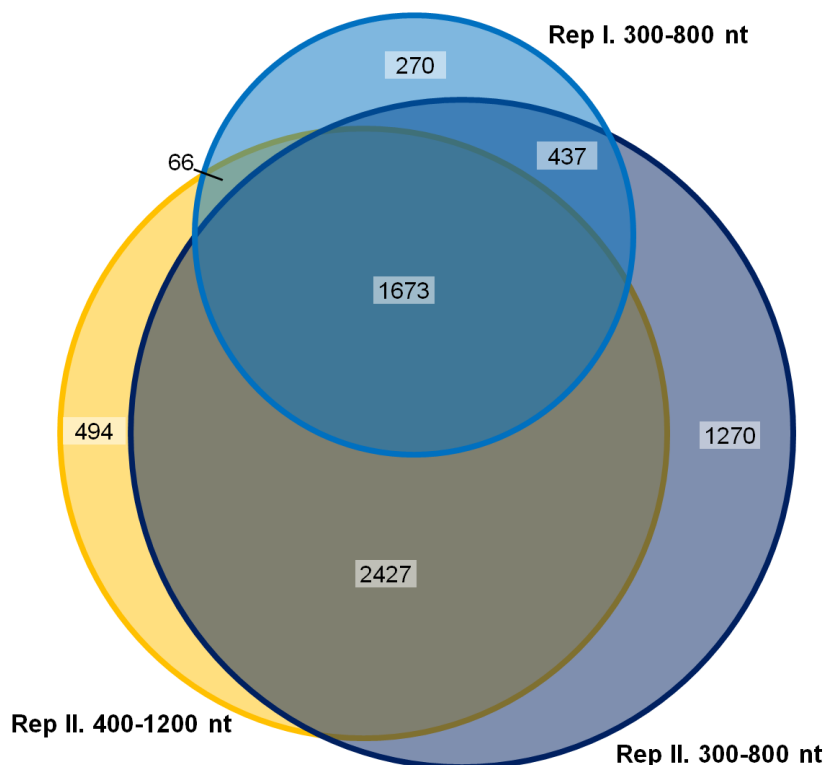


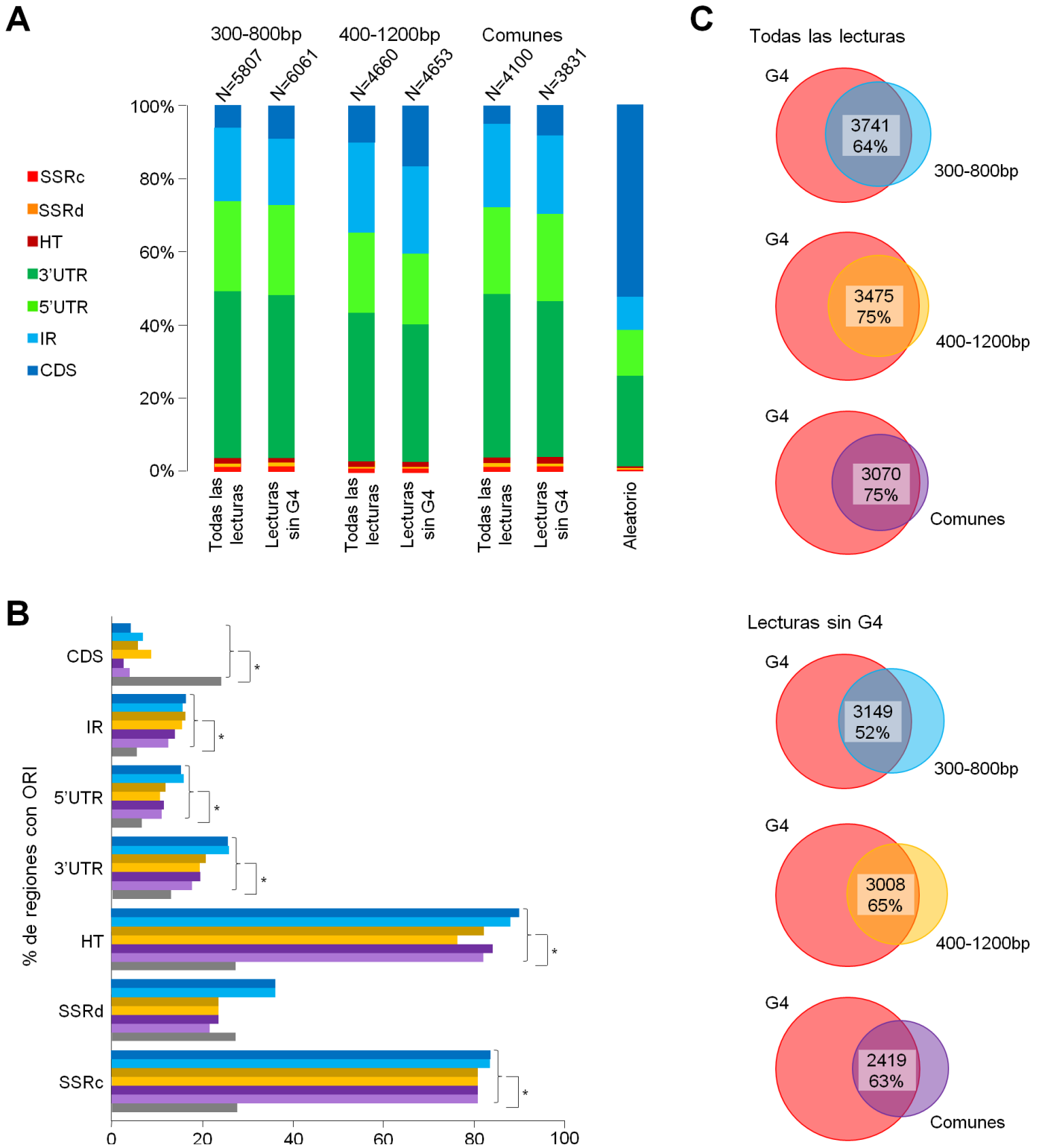
Figura 22: Solapamiento entre las regiones de ORI definidos en los tres experimentos de SNS-seq: Rep I y Rep II 300-800 nt y Rep II 400-1200 nt.

En este caso, los resultados fueron mucho más completos -identificamos una colección de ORI que duplicaba a los de la réplica anterior-. Observamos de nuevo que no había diferencias significativas si el análisis lo realizábamos con todas las lecturas secuenciadas o eliminando aquellas que contenían motivos G4 en sus extremos (**Figura 23A y B**). Además, en esta segunda réplica, el porcentaje de ORI localizados en zonas entre PTU era del 10%, quedando el resto repartido en los límites de las zonas codificantes (**Figura 23A**). Más llamativo aún era comprobar que en este caso el 80-90% de regiones de terminación de transcripción (SSRc y HT) presentaban ORI de manera altamente significativa (**Figura 23B**). También era significativa ($p < 0,0001$) respecto a una distribución aleatoria la presencia de ORI en regiones transcritas pero no codificantes (UTR e IR) así como la baja aparición de ORI en CDS.

Estos resultados nos permiten definir dos clases de ORI en el genoma de *L. major*, aquellos que se localizan en regiones de terminación de la transcripción y aquellos que se localizan en las regiones de *trans-splicing*.

Al examinar la eficiencia de activación de ambas clases de ORI (calculada integrando el número de lecturas normalizadas por el área del mismo), observamos que la correspondiente a los ORI asociados a TTS (SSRc y HT) era significativamente más alta que la de los ORI no asociados a TTS (**Figura 24A**). Además, aunque el estudio se había realizado a escala genómica, quisimos validar estos resultados mediante qPCR en preparaciones independientes de SNS. En la **Figura 24B** reflejamos que tanto la localización como la eficiencia relativa de una representación aleatoria de ORI asociados y no asociados con regiones TTS eran coherentes con los resultados obtenidos a nivel genómico. En conjunto, estas observaciones sugerían que, aunque el grupo de ORI asociados a TTS suponía menos del 4% de los identificados, se utilizaban con una mayor frecuencia en la población de células, concordando con el hecho de que las regiones de TTS estuvieran altamente enriquecidas en ellos.

Figura 23: Caracterización de los ORI definidos en la segunda réplica del estudio en *L. major*. **A:** Distribución genómica de los picos identificados en dos rangos de tamaño de SNS así como los coincidentes en ambas fracciones. Realizamos el análisis con todas las lecturas alineadas o sólo con aquellas que no contuviesen motivos G4 en sus extremos. El histograma de la derecha representa el porcentaje de localización esperada si los ORI se distribuyeran aleatoriamente a lo largo del genoma. **B:** Porcentaje de las regiones genómicas que presentan ORI en los tres casos anteriores así como la comparativa con los esperados en un reparto aleatorio; en azules 300-800 bp, en ocres 400-1200 bp, en violetas los comunes a ambas y en gris, los esperados aleatoriamente ($p < 0,0001$). De cada par de colores, el más oscuro correspondía a los ORI obtenidos con todas las lecturas y el más claro, a los obtenidos filtrando aquellas lecturas que contuvieran motivos G4 en los extremos. **C:** Solapamiento entre los motivos de G4 y los ORI definidos en cada una de las fracciones de intermediarios de replicación y los comunes a ambas.



Finalmente, al realizar la intersección entre los ORI y los motivos G4, observamos que el 64% de ellos solapaban con estos motivos en la fracción de 300-800 nt, incrementándose hasta el 75% en la fracción 400-1.200 nt y en los comunes a ambas fracciones. Este porcentaje de solapamiento, aunque disminuía ligeramente, era también muy elevado cuando considerábamos los ORI redefinidos tras eliminar del análisis las lecturas con motivos G4 en sus extremos (**Figura 23C**).

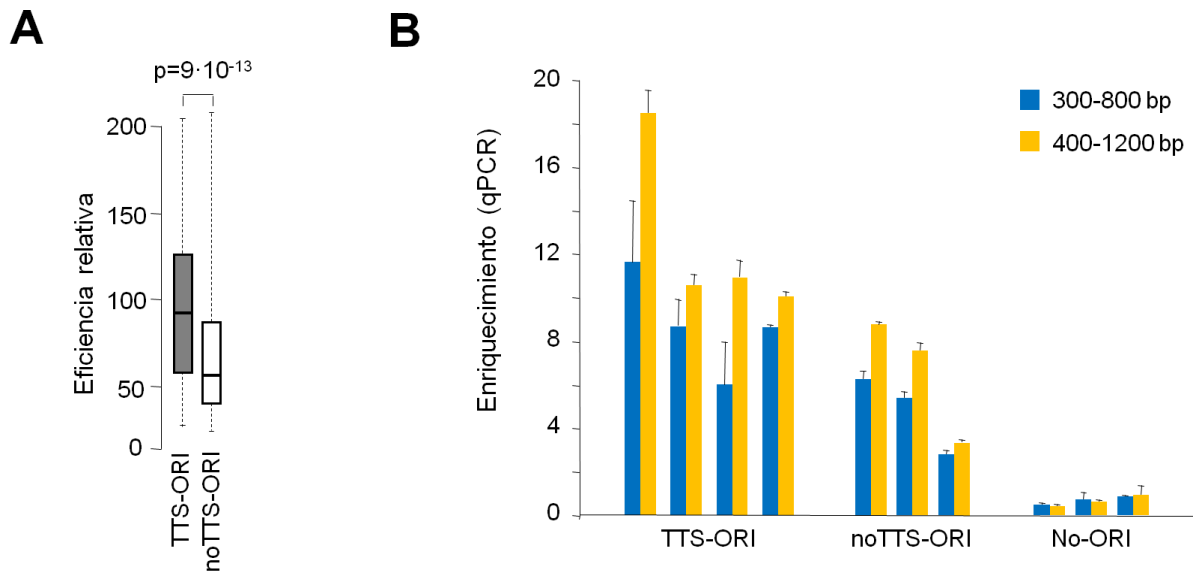


Figura 24: Eficiencia de ambas clases de ORI calculada integrando el número de lecturas acumuladas en cada uno de ellos normalizados por el área del pico que los define **(A)**. Los ORI asociados a TTS presentan una eficiencia significativamente mayor ($p=9 \cdot 10^{-13}$) que los que no están ligados a TTS. **(B)** Validación por qPCR de ejemplos representativos de ambas clases: 4 asociados a TTS, 3 no asociados y 3 regiones negativas en SNS-seq.

Dada la correspondencia entre esta asociación y la descrita en otros sistemas, decidimos analizar en más detalle la distribución de las cadenas nacientes respecto a los motivos G4 en *L. major*. Tal y como se había descrito previamente en células de ratón (Cayrou *et al.*, 2012), encontramos una acumulación de motivos G4 a ambos lados del máximo de SNS de los ORI analizados, aproximadamente a +/- 50-100 bp **(Figura 25A)**. Al tomar como referencia los motivos G4, advertimos una distribución asimétrica de SNS de tamaños crecientes a ambos lados de estas secuencias, según la hebra de DNA en la que se localizasen dichos motivos **(Figura 25B)**. Esta acumulación de SNS asimétrica respecto al motivo G4 es similar a la descrita en células de *D. melanogaster*, lo que sugiere que las estructuras de G4 podrían actuar como barreras frente al avance de las horquillas de replicación *in vivo* (Comoglio *et al.*, 2015). Integrando todos estos resultados, podíamos descartar la posibilidad de que la metodología empleada estuviera introduciendo sesgos significativos hacia un enriquecimiento en falsos intermediarios de replicación en zonas con formación de estructuras de G4.

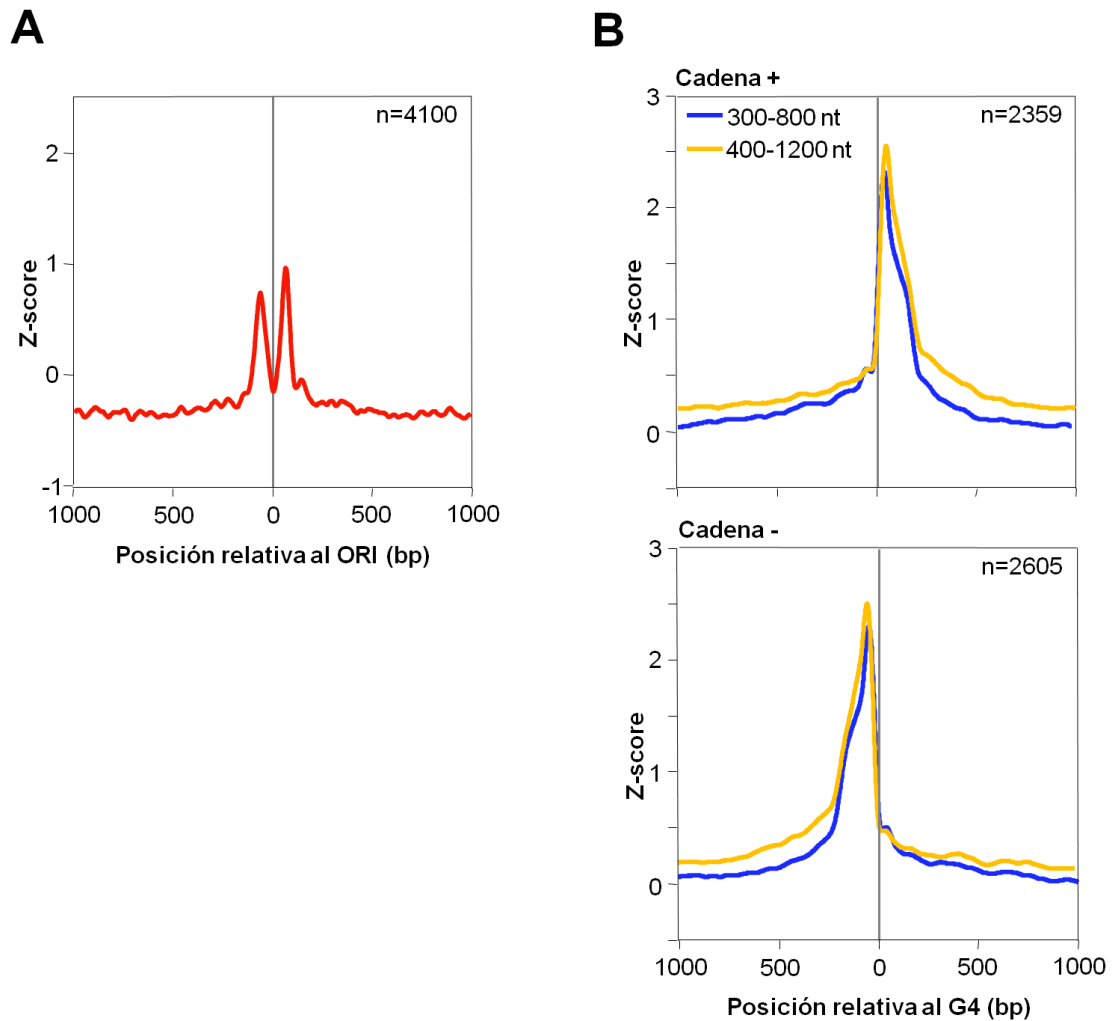
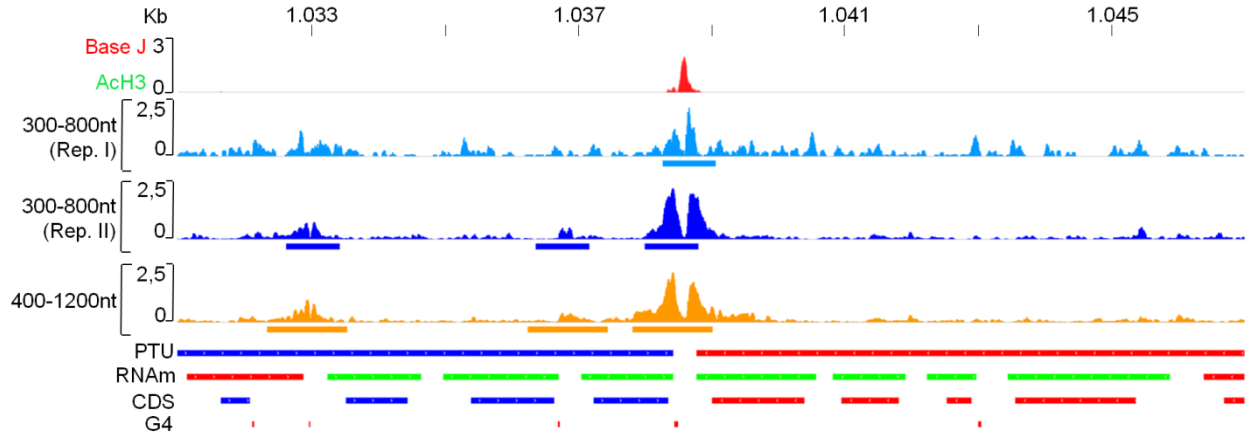


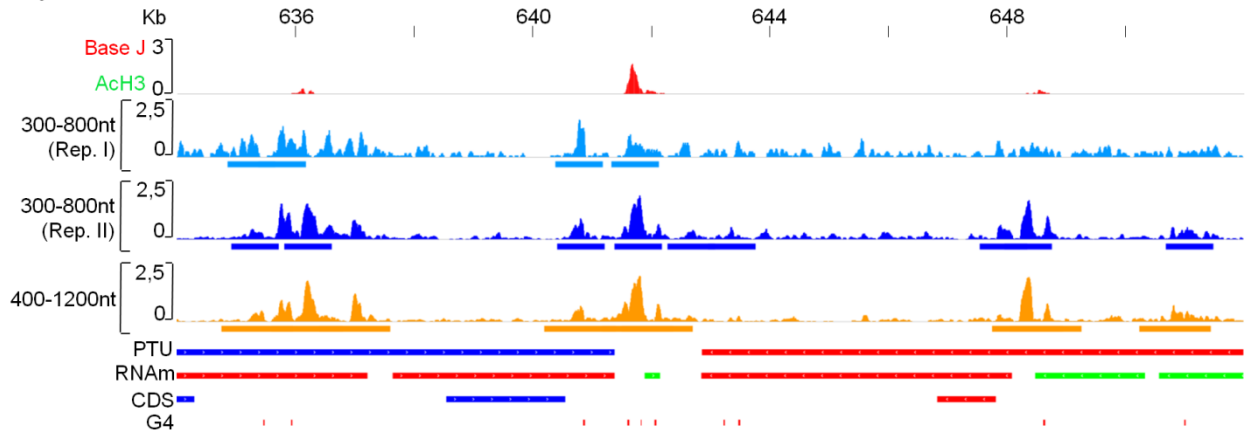
Figura 25: Relación entre los ORI y motivos G4. **(A)** Distribución de motivos G4 respecto al punto máximo de enriquecimiento en SNS de los ORI descritos (n=4100). **(B)** Densidad de SNS de tamaños crecientes de cadenas nacientes (300-800 nt y 400-1200 nt) respecto a los motivos G4 analizados separadamente en función de la hebra de DNA en la que se encuentren.

En la **Figura 26** se representan cuatro segmentos del genoma de *L. major* ilustrando los resultados de la distribución de SNS a lo largo de las distintas regiones analizadas. En los dos primeros paneles se muestran los perfiles de intermediarios de replicación flanqueando regiones SSRc y, en los dos últimos, regiones HT. Además de la localización de ORI en las zonas próximas al TTS, en los cuatro casos observamos enriquecimientos dentro de las PTU que mapean en las regiones 5'UTR- IR-3'UTR, correspondientes a las zonas en las que ocurre el *trans-splicing* sobre el pre-RNA. También se aprecia en la figura la distribución de las SNS respecto a los motivos G4.

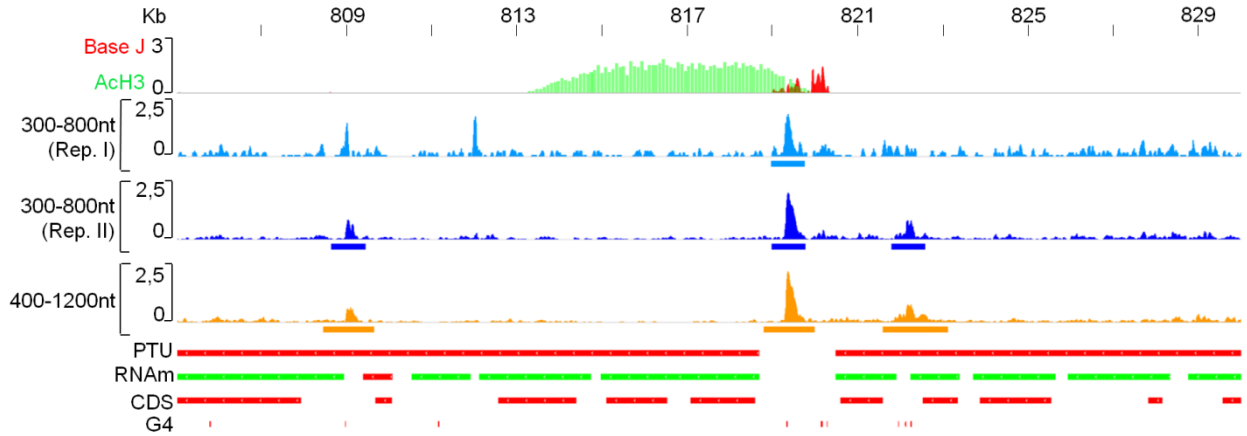
LmjF.28: 1.031.000-1.047.000



LmjF.35: 634.000-652.000



LmjF.29: 805.000-830.000



LmjF.36: 1.103.000-1.128.000

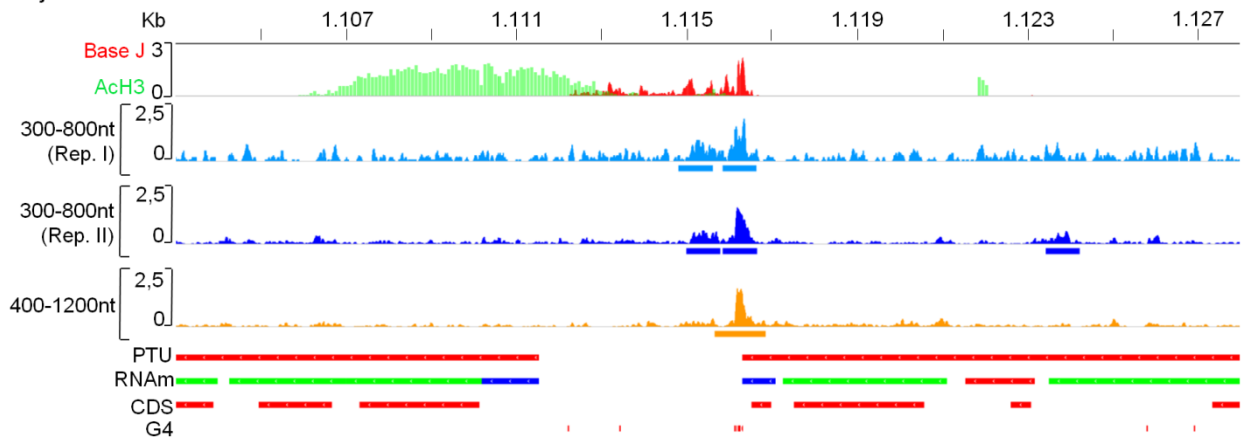


Figura 26: Ejemplos representativos de ORI en cuatro regiones de los cromosomas de *L. major*: dos conteniendo regiones SSRc (primer y segundo panel) y dos conteniendo regiones HT (tercer y cuarto panel). En cada panel se muestran, de la parte superior a la inferior, los datos de localización de la base J en rojo (van Luenen *et al.*, 2012) y de acetilación Ach3 en verde (Thomas *et al.*, 2009); los perfiles de SNS con los picos que definen para cada una de las fracciones del estudio; la localización de las PTU en azul (hebra +) o rojo (hebra -), los RNAm (Rastrojo *et al.*, 2013) en rojo, verde o azul en función de su abundancia (alta, media o baja), los CDS en azul o rojo según el sentido de la hebra codificante y los motivos G4 definidos computacionalmente.

3. Análisis genómico de la organización nucleosomal en *L. major*

Para estudiar la organización de los nucleosomas en los cromosomas de *L. major* generamos mapas genómicos del paisaje nucleosomal mediante MNasa-seq. Para ello, purificamos el DNA nucleosomal de dos muestras biológicas independientes a partir de cultivos en fase exponencial de crecimiento,

Como en el caso de los datos de SNS-seq, en primer lugar comparamos el comportamiento de los alineadores BWA y Bowtie sobre las lecturas de la primera réplica. En este caso, observamos que Bowtie, el alineador utilizado en el mapeo de ORI, fallaba en algunos puntos, dejando zonas sin lecturas alineadas (**Figura 27**). Aunque estos resultados podrían interpretarse como regiones libres de nucleosomas, comprobamos que muchas de ellas quedaban completadas al utilizar el alineador BWA. Así pues, nos decantamos por esta metodología para el alineamiento de las lecturas correspondientes a MNasa-seq en ambas réplicas consiguiendo una profundidad de 76.176.788 y 83.717.687 pares de lecturas alineadas. Este alto grado de profundidad, con una cobertura superior a 350 lecturas por nucleótido, permitía analizar la organización de los nucleosomas en los cromosomas de *L. major* con máxima resolución. Tras este paso, aplicamos el software NUCwave (Quintales *et al.*, 2015) para la definición del perfil de picos y valles previo al análisis completo. Este algoritmo aplica una función de onda sobre la densidad de las lecturas alineadas convirtiéndola en un perfil suavizado y más definido.

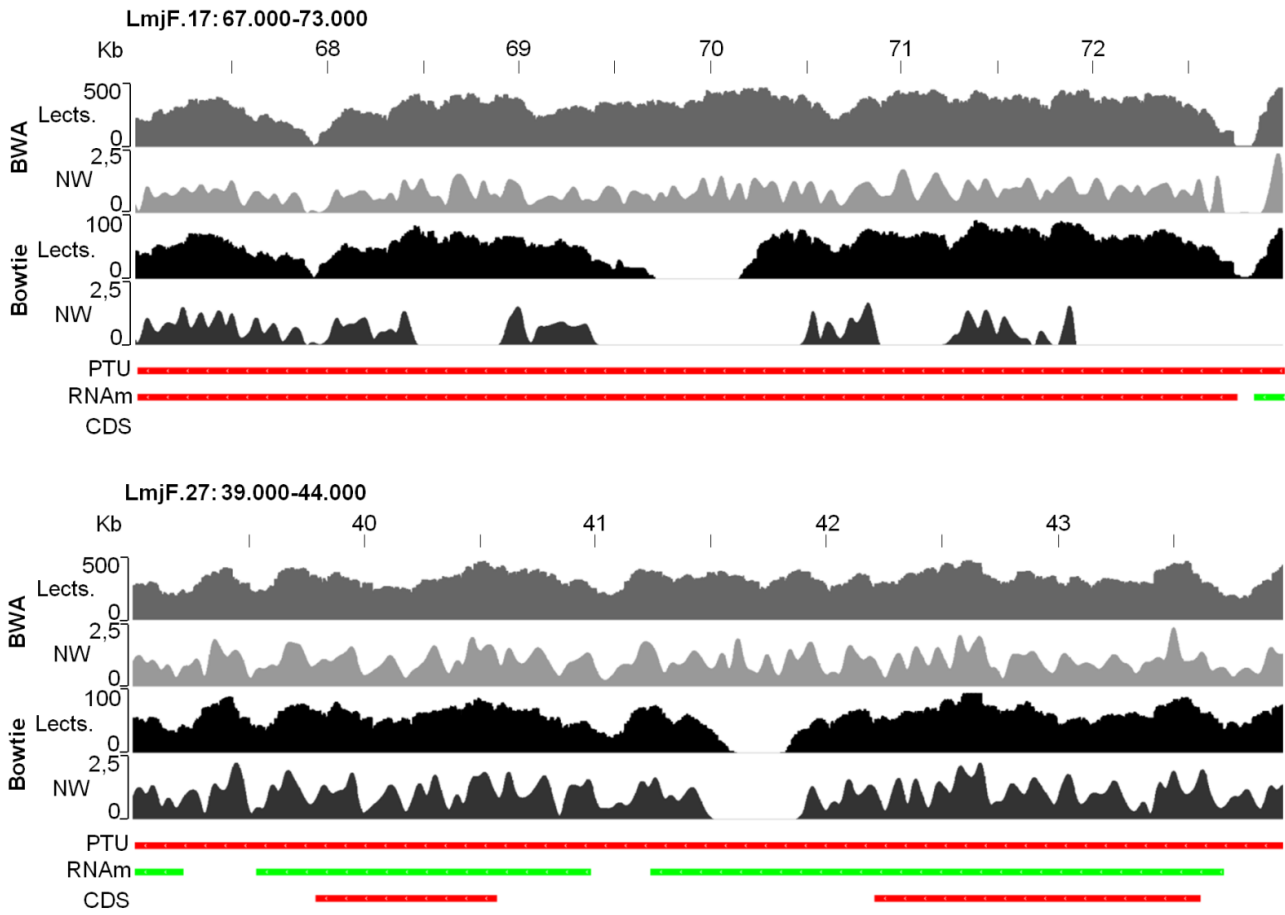


Figura 27: Ejemplos representativos de dos regiones de baja mapeabilidad del genoma de *L. major* ilustrativas de la mejor resolución del alineamiento de lecturas de MNasa-seq con BWA (primera línea) respecto a Bowtie (tercera línea). En la segunda y cuarta línea se muestran los resultados del procesamiento del perfil de lecturas de cada alineador mediante la aplicación del software NUCwave (Quintales *et al.*, 2015). En las líneas inferiores se representan las unidades de transcripción (PTU) en azul (hebra +) o rojo (hebra -), los RNAm (Rastrojo *et al.*, 2013) en rojo, verde o azul en función de su abundancia (alta, media o baja) y los CDS en azul o rojo según el sentido de la hebra codificante.

La correspondencia entre los resultados de las dos réplicas fue muy alta, a pesar del bajo grado de posicionamiento general detectado a lo largo de las PTU (**Figura 28**). Los análisis de perfiles agregados mostraban una menor ocupación nucleosomal respecto al promedio genómico en las zonas de terminación de la transcripción -SSRc y HT- (**Figura 29A**), pero no en las asociadas a los TSS (**Figura 29B**, SSRd). Además, en las regiones HT, en las que la transcripción termina e inicia en la misma hebra, la disminución en la densidad de nucleosomas se asociaba claramente al TTS en lugar de al TSS (**Figura 29B**, HT). Curiosamente, la extensión

de la región con menor cobertura de nucleosomas reflejaba la distribución de la abundancia en base J (**Figura 29A y B**). Estos resultados podrían apuntar hacia algún tipo de relación funcional entre la presencia de la base modificada y la exclusión de nucleosomas en estas zonas. De hecho, se ha sugerido en levaduras que las zonas con menor ocupación nucleosomal asociadas a los TTS podrían jugar un papel activo en el mecanismo de terminación de la transcripción (Fan *et al.*, 2010), que en el caso de *L. major*, podría además estar favorecido por la presencia la base J.

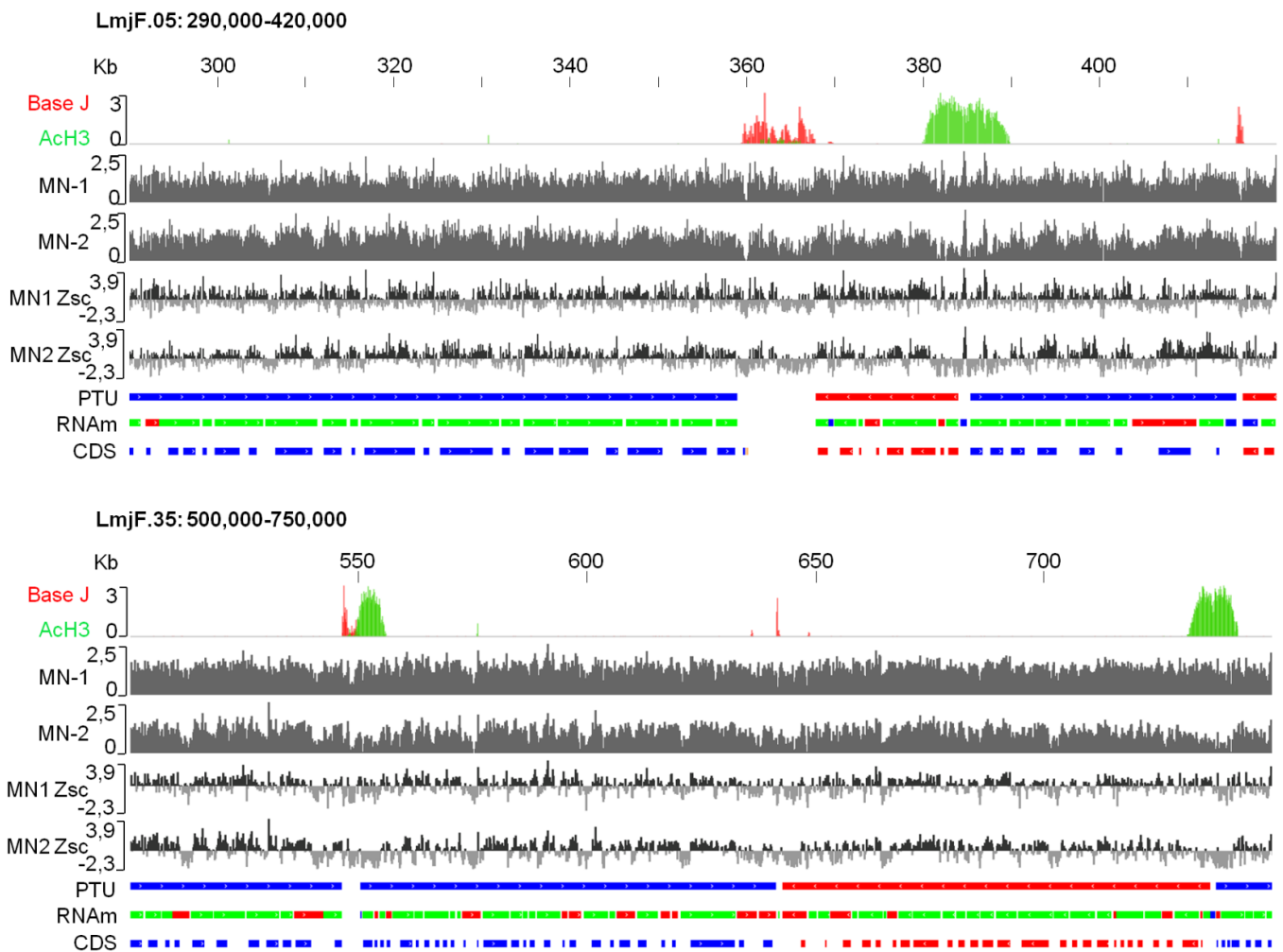


Figura 28: Ejemplos representativos de los resultados de las dos réplicas de MNase-seq de la densidad de nucleosomas (MN-1 y MN-2) y su valor Z-score en dos regiones de los cromosomas (LmjF.05 y LmjF.35) de *L. major*. Se muestran también los datos de localización de la base J en rojo (van Luenen *et al.*, 2012) y de AcH3 en verde (Thomas *et al.*, 2009), ambos en la primera línea. En las líneas inferiores se representan las unidades de transcripción (PTU) en azul (hebra +) o rojo (hebra -), los RNAm (Rastrojo *et al.*, 2013) en rojo, verde o azul en función de su abundancia (alta, media o baja) y los CDS en azul o rojo según el sentido de la hebra codificante.

Como se ha comentado anteriormente, la ocupación nucleosomal fue alta a lo largo de las PTU pero sin un posicionamiento definido (**Figura 28**). Sólo detectamos nucleosomas bien posicionados en las regiones asociadas al comienzo de la transcripción (**Figura 29B**) y en aquellos espacios intergénicos entre el extremo 3'UTR de un pre-RNAm y el 5'UTR del siguiente dentro de la misma PTU (**Figura 29C**). Los nucleosomas posicionados flanqueando al TSS en las posiciones -1 y +1 de regiones SSRd y HT, reproducían una organización similar a la descrita en otros modelos eucariotas (Teif *et al.*, 2012; revisado en Jiang y Pugh, 2009), apoyando la idea del papel de la conformación de la cromatina en la definición del inicio de la transcripción también en las PTU de este organismo.

A la vista de estos resultados, decidimos investigar si el posicionamiento de este nucleosoma en las regiones IR podría estar relacionado con la estabilidad diferencial de los RNAm. Los perfiles agregados de nucleosomas correspondientes a los RNAm analizados según su abundancia relativa mostraron que los 1.086 RNAm más abundantes estaban asociados con NDR mejor definidas (**Figura 29D**). Este resultado, junto con el hecho de que la distancia entre la mayoría de los pre-RNAm individuales de una PTU es de 145 bp, sugería una posible relación de la estructura de la cromatina con el procesamiento transcripcional en este parásito (*trans-splicing*). De hecho, en modelos como *S. cerevisiae* y células de mamíferos, se ha demostrado la importancia del posicionamiento nucleosomal en el proceso de *splicing* (Patrick *et al.*, 2013; Spies *et al.*, 2009). Esto apuntaría a la existencia de un mecanismo similar entre la eliminación de los intrones y la transformación del policistrón transcrito por la RNAPII en los RNAm individuales en *L. major* mediante *trans-splicing*.

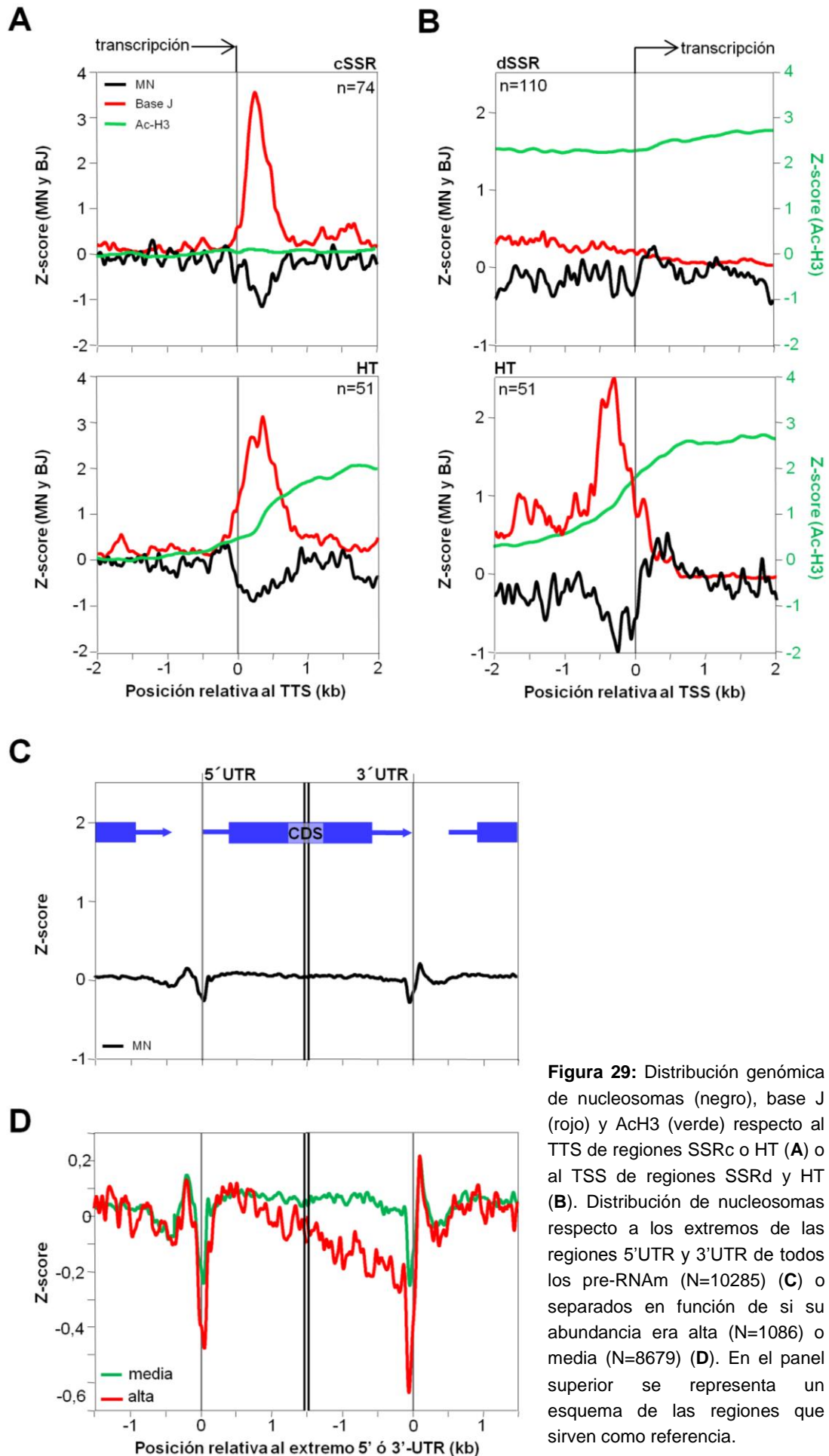


Figura 29: Distribución genómica de nucleosomas (negro), base J (rojo) y AcH3 (verde) respecto al TTS de regiones SSRc o HT (A) o al TSS de regiones SSRd y HT (B). Distribución de nucleosomas respecto a los extremos de las regiones 5'UTR y 3'UTR de todos los pre-RNAm (N=10285) (C) o separados en función de si su abundancia era alta (N=1086) o media (N=8679) (D). En el panel superior se representa un esquema de las regiones que sirven como referencia.

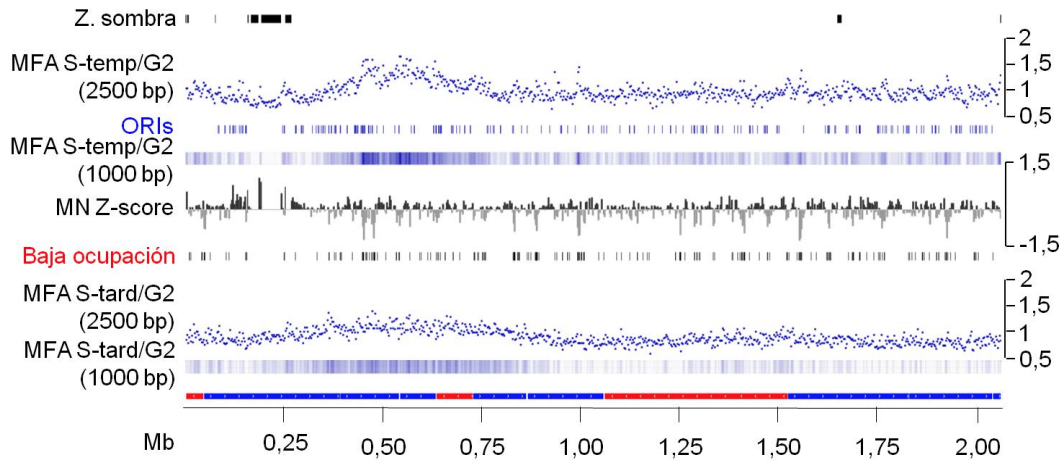
4. Relación entre la estructura de la cromatina y los mecanismos de replicación y transcripción en *L. major*

Un estudio reciente ha descrito, utilizando la técnica de Análisis de Frecuencia de Marcadores (MFA), el panorama replicativo de *L. major*, asignando un solo ORI por cromosoma (Marques *et al.*, 2015). El MFA consiste en la comparación de la profundidad de lecturas de secuenciación derivadas de DNA genómico de células en fase S respecto a células en fase G2. Como las regiones que se replican antes durante la fase S están más representadas en la población que las que se replican después, este tipo de análisis se ha utilizado para estimar la temporalidad de la replicación en varios sistemas, desde *S. cerevisiae* a células circulantes en sangre de individuos sanos (Müller *et al.*, 2014; Koren *et al.*, 2014). En el caso de *S. cerevisiae*, además, como se conoce la localización y actividad de los ORI, estos experimentos han permitido determinar el tiempo de activación a cada uno de ellos (Raghuraman *et al.*, 2001; Müller *et al.*, 2014). La resolución de MFA en organismos de genomas más grandes no es lo suficientemente alta como para asignar las regiones de replicación temprana a ORI individuales, por lo que estos dominios se han interpretado como regiones de estructura de cromatina similar que comprenden numerosos ORI que se activan en tiempos similares durante la fase S (revisado en Marks *et al.*, 2016). Para comparar nuestros resultados de localización de ORI mediante SNS-seq con los del paisaje replicativo inferido mediante MFA, incrementamos la resolución de las ventanas del análisis (2500 bp en el trabajo original, Marques *et al.*, 2015) por una más cercana a la resolución de nuestros ORI (ventanas solapantes de 1.000 bp). Al examinar ambos grupos de datos a lo largo de los cromosomas de *L. major* encontramos una correlación significativa ($p < 10^{-3}$) entre la densidad de ORI determinada por SNS-Seq y las regiones de alto valor de MFA (**Figura 30 A y B**). Estos resultados sugerían que los análisis de MFA en *L. major* realmente mostraban información temporal y que de ellos no podría extrapolarse directamente la localización de los ORI.

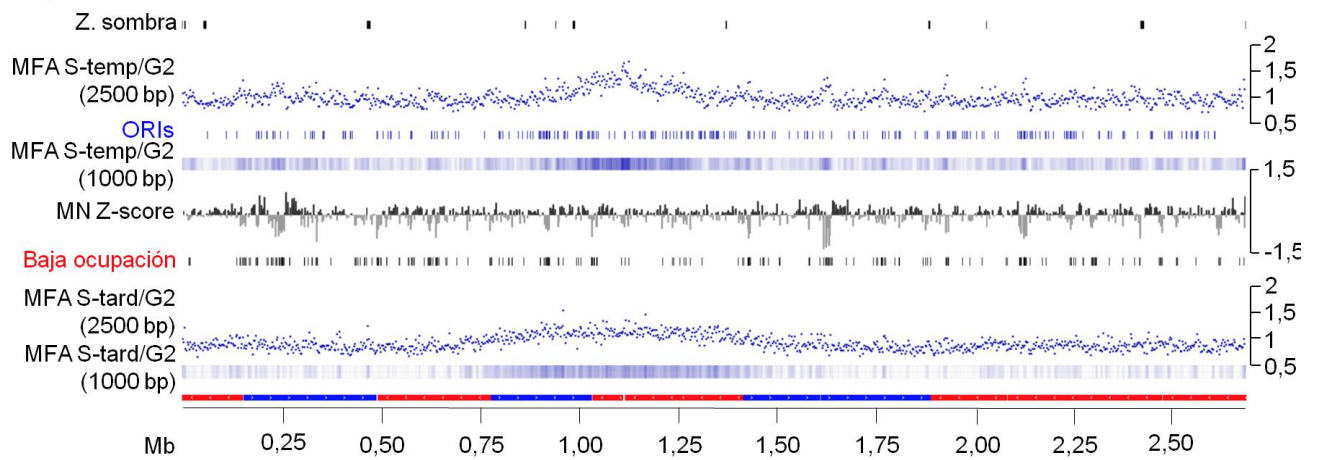
Figura 30: Las regiones de replicación temprana contienen una mayor densidad en ORI de regiones de baja ocupación nucleosomal. **(A)** Representación de los dos cromosomas de *L. major* de mayor longitud en los que se muestra desde la parte superior a la inferior: (i) MFA de fase S-temprana/G2 analizado en ventanas de 2500 bp (Marques *et al.*, 2015), (ii) ORI identificados mediante SNS-seq, (iii) MFA de fase S-temprana/G2 analizado en ventanas de 1000 bp, (iv) ocupación nucleosomal (*Z-score*), (v) zonas de baja ocupación nucleosomal, (vi) MFA de fases S-tardía/G2 analizado en ventanas de 2500 bp (Marques *et al.*, 2015) y (vii) MFA de fases S-tardía/G2 analizado en ventanas de 1000 bp, respectivamente. Los rectángulos negros de la parte superior representan las zonas de sombra excluidas del estudio. **(B)** Significación de la correlación entre la densidad de ORI y las zonas de

A

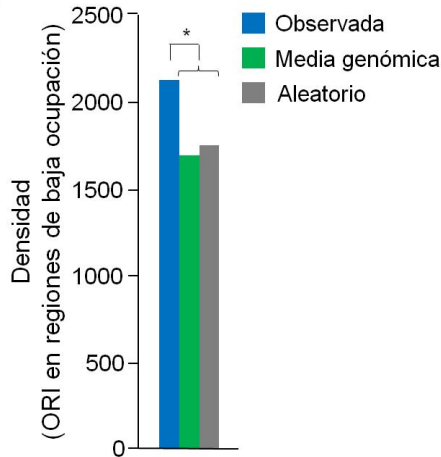
LmjF.35:



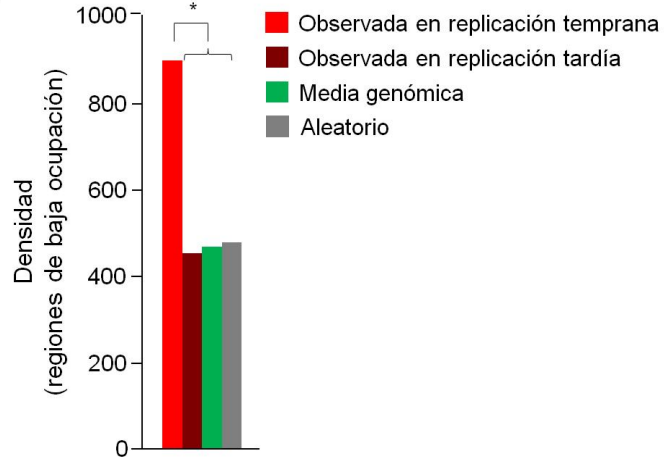
LmjF.36:



B



C



replicación temprana ($p < 0,001$). Los histogramas representan la densidad de ORI observada en zonas de replicación temprana (azul); densidad media de ORI en el genoma (verde) y densidad de fragmentos genómicos aleatorios en zonas de replicación temprana (gris). (C) Significación de la correlación entre la densidad de regiones de baja ocupación nucleosomal y las zonas de replicación temprana ($p < 0,001$) o tardía ($p = 0,84$). Los histogramas representan la densidad de zonas de baja ocupación nucleosomal en zonas de replicación temprana (rojo) o tardía (granate); densidad media de zonas de baja ocupación nucleosomal en el genoma (verde) y densidad de fragmentos genómicos aleatorios (gris).

Tanto en células humanas (Gindin *et al.*, 2014) como en *D. melanogaster* (Eaton *et al.*, 2011) se ha descrito que las zonas de replicación temprana presentan una mayor accesibilidad al DNA. De hecho, en el trabajo de Gindin *et al.* (2014) elaboran un modelo de cinética de replicación que es capaz de recapitular el programa temporal de la replicación en células humanas simplemente teniendo en cuenta la distribución de los sitios de hipersensibilidad a DNAsa I. Para analizar si la accesibilidad al DNA podría predecir las regiones de replicación temprana en *L. major*, estimamos las regiones más expuestas del genoma a partir de nuestros resultados de los Z-score de los experimentos de MNasa-seq (**Figura 30A**). Encontramos una correlación significativa ($p < 10^{-3}$) entre la densidad de las regiones con baja ocupación nucleosomal y las regiones de alto MFA (**Figura 30C**, panel izquierdo). Esta correlación desaparecía ($p = 0,84$) cuando se comparaban las zonas de baja ocupación nucleosomal con los valores del MFA de células en fase S tardía (**Figura 30C**, panel derecho). Estos resultados indican que, de modo similar a lo descrito en otros sistemas, la estructura de la cromatina en *L. major* es más accesible en las zonas de replicación temprana.

A continuación analizamos en detalle la relación entre los ORI y los nucleosomas en los distintos compartimentos genómicos. En los ORI asociados a TTS encontramos una clara correlación entre el enriquecimiento en SNS y la baja ocupación nucleosomal (**Figura 31A**). Del mismo modo que intuíamos en los ejemplos representativos de la **Figura 26**, el máximo de la señal de SNS parecía flanquear al máximo de localización de la base J, aunque este efecto era más claro en las regiones HT (que sólo poseen un TTS) que en las SSRc, ya que al contener dos TTS, el solapamiento de las densidades dificultaba la distinción. Esta asociación entre la abundancia de SNS y la baja ocupación nucleosomal también se observaba al analizar los ORI localizados en las regiones genómicas en las que ocurría el *trans-splicing*, (3'UTR-IR-5'UTR) (**Figura 31B**). Hay que tener en cuenta que en estas regiones el enriquecimiento en SNS era menor debido a que los perfiles agregados representan Z-scores y sólo el 20% de todas ellas contenían ORI. Por último, analizamos la distribución de los nucleosomas alrededor del punto de máximo enriquecimiento en SNS de los ORI asociados y no asociados a TTS (**Figura 31C**). El resultado corroboraba la co-localización entre las zonas de baja ocupación nucleosomal con los ORI. En el caso de los TTS-ORI (panel inferior), la coincidencia de la región de baja ocupación de nucleosomas con el máximo de la señal de SNS era más clara, probablemente debido a la menor ocupación nucleosomal detectada en las regiones de TTS.

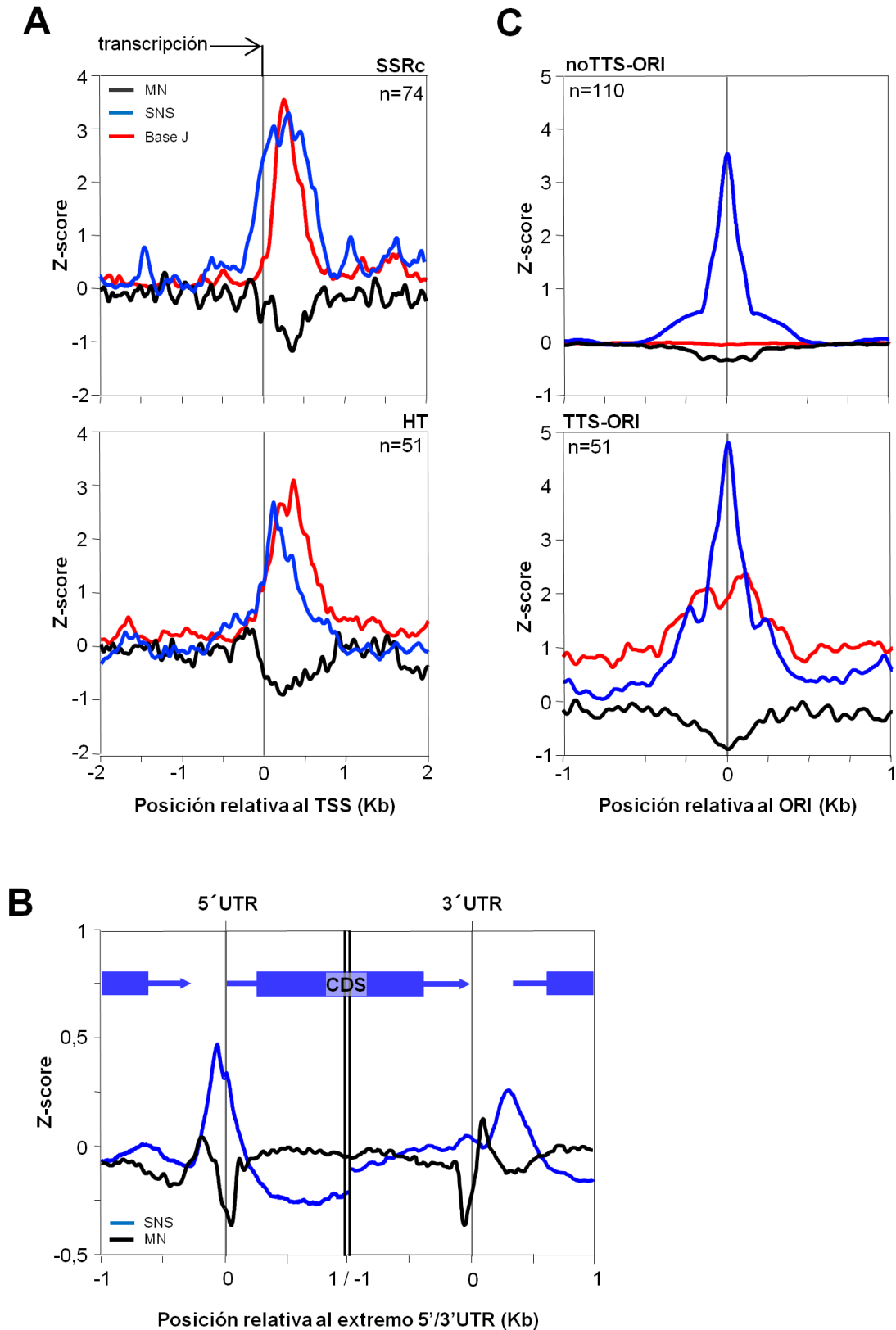


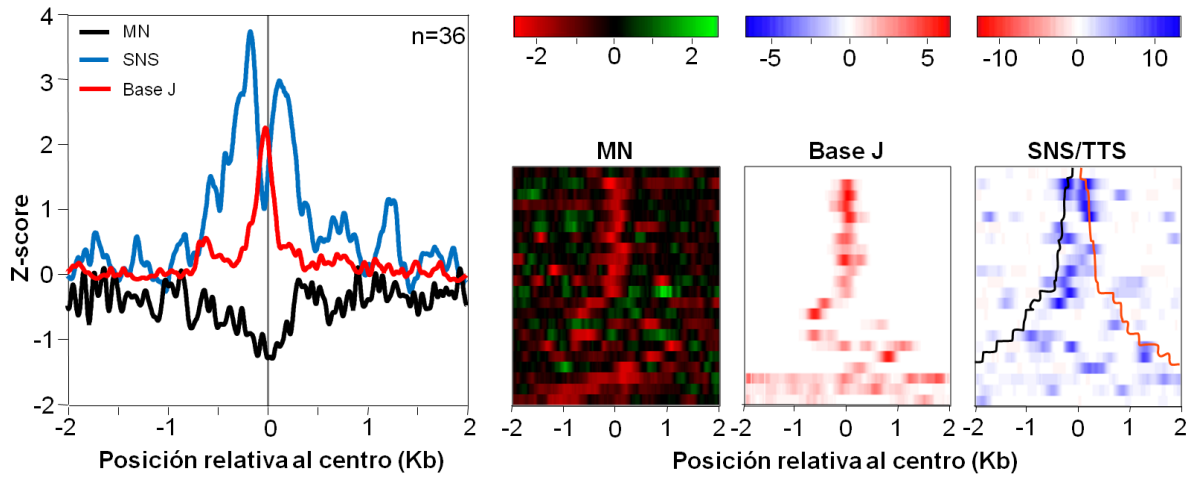
Figura 31: Distribución genómica de nucleosomas (negro) y SNS (azul) en *L. major* respecto a los TTS (A); a los extremos de las regiones 5'UTR y 3'UTR de los pre-rNAM (B) o al punto de máximo enriquecimiento de SNS de los ORI divididos en TTS-ORI o noTTS-ORI (C).

Dada la correlación existente entre el inicio de la replicación y una menor ocupación nucleosomal, decidimos profundizar en el análisis de la relación entre ambos parámetros con el proceso de transcripción. Los resultados del mapeo de ORI indicaban que la iniciación de la replicación en *L. major* ocurría en los sitios de pausa y terminación de la transcripción de la RNAPII: regiones de *trans-splicing* y TTS, respectivamente. Para analizar el efecto de la direccionalidad de la transcripción, tanto en la ocupación nucleosomal, como en la abundancia de SNS, estratificamos las regiones SSRc y HT de menor a mayor de acuerdo con la distancia entre ambos TTS (**Figura 32A**, SSRc) o entre el TTS en posición 5' y el TSS contiguo (**Figura 32B**, HT) y representamos la distribución de la ocupación nucleosomal, la base J y el enriquecimiento en SNS. Estos análisis confirmaban la menor ocupación nucleosomal de las regiones no transcritas respecto a las PTU, así como que la transición entre alta y baja ocupación coincidía con la posición del TTS, apareciendo en los puntos de menor ocupación nucleosomal un fuerte enriquecimiento en base J. Esto podría deberse al hecho de que esta base modificada excluya a los nucleosomas, lo que también podría contribuir a su papel en la terminación de la transcripción (van Luenen *et al.*, 2012). Tanto la disminución de la ocupación nucleosomal como el aumento de base J en el centro de las SSRc y HT eran significativos respecto a las PTU que convergían o finalizaban en ellos, donde el nivel de ambos volvía a situarse en el promedio genómico. Por otra parte, la distribución de SNS revelaba la existencia de dos máximos inmediatamente adyacentes a cada uno de los TTS en las regiones SSRc (**Figura 32A**) o de un único máximo adyacente al único TTS en el caso de las regiones HT (**Figura 32B**). En conjunto, nuestros resultados sugerían que la dirección de elongación de la transcripción de la RNAPII influye en la localización de los ORI en el genoma de *L. major*.

Figura 32: Distribución de nucleosomas, SNS y base J a lo largo de 4 Kb respecto al centro de 36 regiones SSRc (**A**) o 72 regiones HT (**B**). Las regiones aparecen ordenadas de arriba abajo según la distancia entre ambos TTS (SSRc) o entre un TTS y su TSS contiguo (HT). En el panel izquierdo representamos el perfil agregado de las densidades de las tres señales. En los mapas de color, separados por categoría, se muestran los *Z-score* de cada una de las señales.

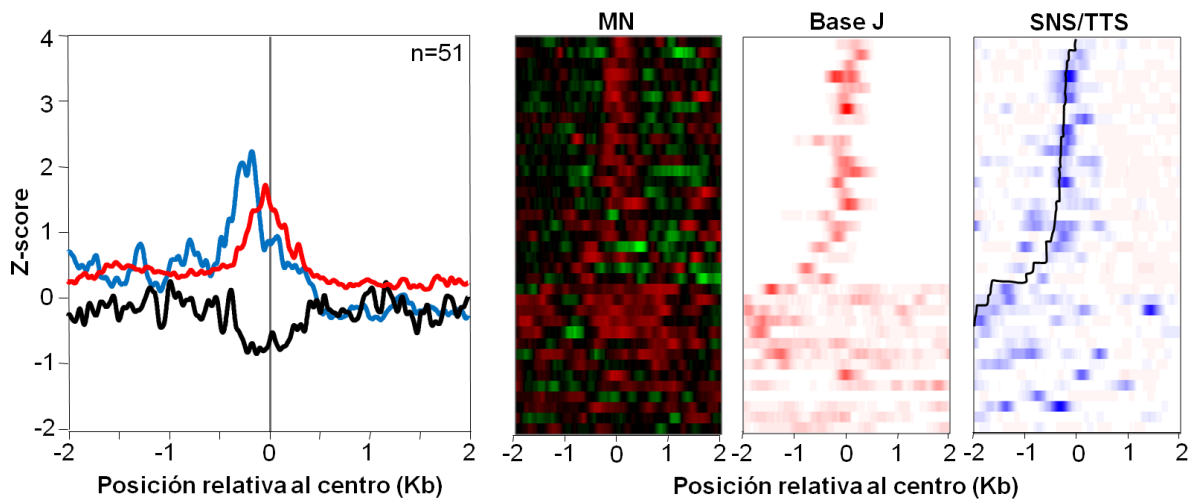
A

SSRc



B

HT



De las tres funciones vitales, la reproducción ocupa un nivel jerárquico superior respecto a la relación y a la nutrición, sirviéndose de éstas en último caso, para sus propios fines. A nivel celular esto ocurre de modo similar: cada célula se relaciona con las de su entorno -ya sea con las que conforma un organismo pluricelular mayor o con las de la población de unicelulares a la que pertenece- y se nutre de los compuestos necesarios para mantener a su maquinaria celular con el fin último de, por medio de la reproducción, dar lugar a dos células hijas. Profundizando a una escala bioquímica, este proceso requiere la duplicación de los componentes celulares, entre ellos, el de mayor importancia, la información codificada en el DNA. ¿Es la replicación del DNA el fin último de la función de reproducción que, a su vez, mantiene un lugar privilegiado frente a las demás? Aun siendo una aproximación más ontológica que científica, si así fuese, el inicio de la replicación, su regulación, así como el reconocimiento de los lugares apropiados para su activación conformaría uno de los pilares básicos de la vida justificando por sí mismo la necesidad de ahondar en el conocimiento de este campo.

Para profundizar en la comprensión del papel que juegan diversos factores sobre la especificación de una región genómica como posible ORI, hemos llevado a cabo dos enfoques claramente diferenciados. Por una parte, una aproximación a escala genómica en los sistemas de células de ratón, humanas o promastigotes de *L. major* integrando los datos disponibles en la literatura o generados por nosotros respecto a, entre otros, motivos de la secuencia de DNA -contenido en G/C, motivos G4 y formación de *R-loops*-, modificación de histonas -presencia de las variantes H3.3 y H2A.Z o acetilación de H3 en *L. major*- o bases modificadas -base J característica de tripanosomátidos- junto con el paisaje nucleosomal, la actividad transcripcional y el perfil del inicio de la replicación del DNA generado a partir de SNS o la localización del complejo ORC. Por otro lado, hemos realizado un análisis de alta resolución de la densidad de SNS y ocupación nucleosomal en promotores-ORI específicos. Los resultados obtenidos contribuyen a alejar la idea de que las regiones por las que se inicia la síntesis del DNA se especifiquen por una causa única -como la secuencia consenso característica de las ARS de *S. cerevisiae*. Por el contrario, los datos sugieren que la probabilidad de que una región funcione como ORI depende de una compleja interrelación entre distintos factores.

Durante su ciclo vital, una célula tiene que mantener la integridad de sus componentes y procesos. Alrededor del DNA ocurren simultáneamente, además de su replicación, numerosos procesos como la transcripción, el desensamblaje y remodelado de la cromatina, la reparación de errores, etc. En cualquier caso, como la

duplicación de la información genética para su posterior transmisión a la descendencia es básica en su éxito vital, cada célula debe, necesariamente, llevar a cabo este proceso de manera que no interfiera negativamente con los demás. Es plausible que las células, a lo largo de la evolución, hayan organizado su genoma y los procesos fundamentales que ocurren en él, de modo que se minimicen las interacciones nocivas para su supervivencia. En este sentido, los diversos factores que contribuyen positivamente al inicio de la replicación del DNA, podrían generar un microambiente en la doble hélice favorable al reclutamiento de los complejos de licenciamiento que determinan el que una región pueda ser activada como ORI, bien especificando la unión de los complejos ORC, o bien incrementando la probabilidad de que esa región sea seleccionada para su activación e inmediata iniciación de la replicación en un escenario estocástico. Este enfoque probabilístico de la capacidad de una región para activarse como ORI, además de explicar las interacciones entre los diferentes agentes y procesos nucleares, podría justificar las divergencias en la cartografía de orígenes a nivel genómico que muestran los diferentes estudios en función de su metodología y sensibilidad (**Figura 33**). La hipótesis de esta naturaleza oportunista del inicio de la replicación ha sido expuesta por diversos grupos, incluido el nuestro (Antequera, 2004; Mechali, 2010; Sequeira-Mendes y Gómez, 2012) y a lo largo de esta discusión, profundizaremos en ella y en su causalidad respecto al resto de características estudiadas.

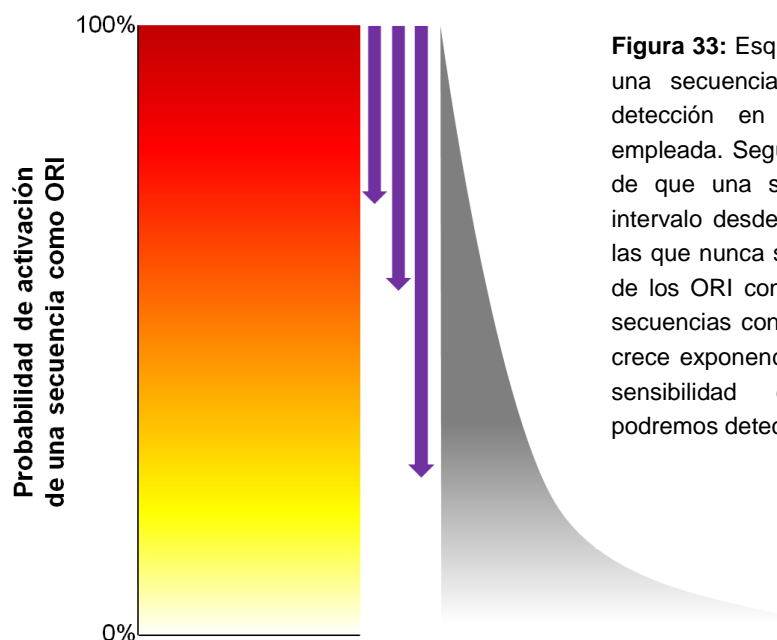


Figura 33: Esquema de la probabilidad de que una secuencia actúe como ORI y de su detección en función de la metodología empleada. Según este modelo, la probabilidad de que una secuencia se active, tiene un intervalo desde el 0% -propio de regiones en las que nunca se inicia la replicación- al 100% de los ORI constitutivos. Como el número de secuencias con baja probabilidad de activarse crece exponencialmente (curva gris), según la sensibilidad del experimento (flechas), podremos detectar un número variable de ORI.

1. Sesgos y motivos en la secuencia del DNA que podrían favorecer la iniciación de la replicación

La relación entre el inicio de la replicación y el mayor contenido en G/C de la secuencia de DNA ha sido apuntada en múltiples estudios realizados en sistemas animales o plantas, como ya se ha desarrollado en la Introducción. Con el aumento del número de ejemplos estudiados gracias a la aplicación de técnicas genómicas al mapeo del perfil replicativo, se ha puesto de manifiesto que los ORI más eficientes se encuentran asociados a promotores de CGI (Sequeira-Mendes *et al.*, 2009; Cayrou *et al.*, 2011), así como a secuencias con potencial de formar estructuras de G4 (Cayrou *et al.*, 2012; Besnard *et al.*, 2012; Picard *et al.*, 2014). También en regiones promotoras asociadas a CGI se ha mapeado la localización de híbridos DNA:RNA o *R-loops*, precisamente en la región inmediatamente posterior al TSS que presenta la asimetría en bases GC-skew (Ginno *et al.*, 2012; 2013; Sanz *et al.*, 2016). En conjunto estos datos sugieren que las regiones preferentes de inicio de la replicación en células de mamíferos tienen en común la riqueza en G/C y, como consecuencia, el enriquecimiento en motivos G4 y en *R-loops*.

Existe cierta controversia metodológica en el campo de la replicación respecto al mapeo genómico de los sitios de síntesis del DNA dependiendo de la metodología utilizada (revisado en MacAlpine y Prioleau, 2016). En particular, se ha argumentado que el enriquecimiento enzimático en SNS mediante la digestión de la enzima λ -exonucleasa podría presentar un sesgo hacia la sobreestimación de secuencias ricas en G/C y/o no digerir completamente las estructuras de G4 (Fouk *et al.*, 2015). Sin embargo, experimentos pioneros de purificación de intermediarios de replicación mediante marcaje e inmunoprecipitación con BrdU, habían demostrado previamente que las CGI funcionan como ORI (Delgado *et al.*, 1998). En este trabajo hemos comprobado mediante distintas aproximaciones que la realización de tratamientos seriados de fosforilación y digestión con λ -exonucleasa minimiza los enriquecimientos no deseados (**Figura 7** y Comoglio *et al.*, 2015). Como control adicional, en los experimentos de mapeo de SNS en *L. major* hemos realizado en paralelo los análisis computacionales con el total de lecturas secuenciadas así como eliminando aquellas que pudieran formar estructuras de G4 en sus extremos, no encontrando diferencias significativas en la distribución genómica de los ORI (**Figuras 21 y 23**).

El hecho de que los motivos G4 sean definidos mediante su secuencia, también provoca disparidad de criterios al asociar su presencia en los ORI. Por ejemplo, los ORI definidos en células humanas IMR-90 solapan en un 67% con

motivos G4 con un bucle máximo de 7 nt pero superan el 91% si esta distancia se extiende a 15 nt (Besnard *et al.*, 2012). La longitud máxima de los bucles en los motivos G4 es un parámetro a tener en cuenta puesto que diversos estudios asignan una baja estabilidad termodinámica a las estructuras con bucles mayores de 7 nt (Hazel *et al.*, 2004; Bugaut y Balasubramanian, 2008). No obstante, experimentos *in vitro* ponen de manifiesto que la distinta naturaleza del catión que estabiliza la estructura puede permitir la formación de bucles por encima del límite de 7 nt (Guédin *et al.*, 2010). Es más, en un estudio reciente que aísla conformaciones de G4 *in vivo* acoplado a secuenciación masiva (G4-seq), describe la aparición de algunas de estas estructuras que no siguen el motivo canónico o comprenden bucles de longitudes superiores a 7 nt (Chambers *et al.*, 2015). En cualquier caso, en nuestros análisis, aun mantenido una posición conservadora al definir como motivos G4 aquellos que contienen bucles de 7 nt, hemos encontrado una clara correlación entre estos y la localización de los ORI en *L. major* (**Figuras 21, 23 y 25**), extendiendo a este organismo las observaciones previas hechas en células de ratón, humanas o de *D. melanogaster* (Cayrou *et al.*, 2012; Besnard *et al.*, 2012; Picard *et al.*, 2014; Comoglio *et al.*, 2015). Por otra parte, los análisis detallados de la localización de motivos G4, *R-loops*, *GC-skew*, unión de la proteína ORC1 y abundancia de SNS en células humanas revelan la íntima relación posicional de estos factores (**Figuras 13 y 14**). Aunque los análisis no permiten establecer causalidad entre estas asociaciones, son compatibles con estudios *in vitro* que demuestran la afinidad incrementada de la proteína ORC1 por DNA de cadena sencilla (ssDNA) o RNA respecto al DNA de doble cadena (dsDNA) (Hoshina *et al.*, 2013), o que la modificación de la secuencia canónica del motivo G4 asociado al ORI de la β -globina en células DT40 minimiza la actividad replicativa del mismo (Valton *et al.*, 2014 y revisado en Valton y Prioleau, 2016). Más complicado resulta asignar el papel funcional que pueden tener los *R-loops* y el *GC-skew* que llevan asociado. No obstante, existen evidencias de que algunos tipos de RNA pueden estabilizar o favorecer la iniciación de la replicación. Por ejemplo, una categoría de RNA no codificante (ncRNA) denominados YRNA, son esenciales para mantener la replicación en extractos celulares humanos (Christov *et al.*, 2006). Por otro lado, en la replicación del virus Epstein-Barr, patógeno de células humanas, el reclutamiento de ORC hacia dos motivos del DNA viral conocidos como LR1 y LR2 está mediado por RNA (Norseen *et al.*, 2008). Estos RNA presentan a su vez motivos G4 y la inhibición química de los mismos impide la replicación vírica así como también disminuye la proliferación de las células hospedadoras (Norseen *et al.*, 2009).

Resulta interesante especular con la posibilidad de que la naturaleza de los *R-loops* -en el que la molécula de RNA anilla con la hebra molde del DNA desplazando a la otra hebra en forma de ssDNA- favorezca en esta última la formación de estructuras de G4 (**Figura 34**). En ese caso, la generación de un híbrido DNA:RNA podría tener como consecuencia el que una secuencia con potencial de establecer un G4, efectivamente lo formase. De hecho, como se ha comentado anteriormente, la afinidad de ORC1 *in vitro* es mayor por secuencias con este tipo de motivos (Hoshina *et al.*, 2013). Otra sugerente explicación, en el sentido inverso, sería que una molécula de RNA pudiera estabilizar la hebra de DNA complementaria cuando aparece una estructura de tipo G4. De hecho, la estabilización de estas estructuras producida por ncRNA se ha observado en la variación antigénica de los *pili* de *Neisseria gonorrhoeae* (Cahoon y Seifert, 2013).

Por tanto, los motivos en la secuencia del DNA podrían ser uno de los factores que contribuyen a generar una determinada topología del DNA a través de la formación de *R-loops* y/o estructuras de G4, lo que podría incrementar la afinidad de unión de los componentes implicados en el reconocimiento de los ORI y/o en su activación.

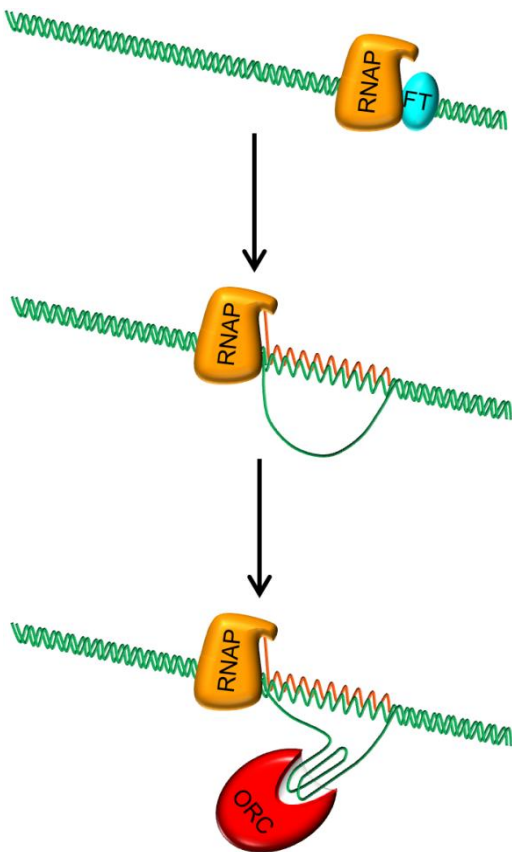


Figura 34: Modelo de un posible escenario en el que las estructuras de *R-loops* y/o G4 favorecen el reconocimiento de un ORI en las inmediaciones de un promotor: cuando la RNAPII comienza la transcripción dejando de cadena sencilla la hebra no molde de DNA, si ésta tiene presencia de motivos G4, la probabilidad de que forme esta estructura será mayor. Además, está documentada la afinidad *in vitro* de ORC por este tipo de estructuras secundarias (Hoshina *et al.*, 2013).

2. El paisaje nucleosomal refleja la probabilidad de iniciación de la replicación del DNA

La accesibilidad del DNA es uno de los factores con mayor poder predictivo para identificar regiones de replicación temprana y, por tanto, que contienen mayor densidad de ORI activos en células humanas (Gindin *et al.*, 2014; Miotto *et al.*, 2016). Nuestros resultados corroboran esta observación en el genoma de *L. major*, ya que las zonas de replicación temprana se corresponden con aquellas de menor ocupación nucleosomal y mayor densidad de ORI (**Figura 30**). Del mismo modo, cada uno de los ORI identificados en este organismo, independientemente de su asociación o no a regiones de terminación de la transcripción, presentan una menor ocupación nucleosomal que sus regiones flanqueantes (**Figuras 29-32**). La misma conclusión puede extraerse de los resultados obtenidos en células de mamíferos (**Figura 7**), aunque análisis detallados de sensibilidad a MNasa indican que esta menor ocupación nucleosomal se debe a la presencia de nucleosomas lábiles presentes en la posición del TSS de promotores activos (**Figura 9**). Estos resultados sugieren un alto recambio de nucleosomas en los promotores activos promovidos por la inestabilidad de la partícula nucleosomal compuesta por las variantes de histonas H2A.Z y H3.3 (Jin *et al.*, 2009). De hecho, la localización de ORC1 en los promotores-ORI más eficientes en células HeLa ocurre exactamente en la posición ocupada por un nucleosoma lábil (**Figura 13**), indicando que ORC podría reclutarse de manera preferencial en promotores activos debido a que presentan una arquitectura nucleosomal más accesible (**Figura 35**). Estos resultados se han confirmado en estudios posteriores en células ES de ratón (Cayrou *et al.*, 2015).

Los análisis de alta resolución de los sitios de inicio de la síntesis del DNA en ORI eficientes han mostrado que estos ocurren en regiones de nucleosomas posicionados, independientemente de si son lábiles o estables (**Figuras 9-11**). Estos resultados ilustran la variabilidad de conformaciones de cromatina en las que puede ocurrir la iniciación de la replicación y sugieren que las variaciones en eficiencia de los ORI podrían deberse a la mayor probabilidad de reclutamiento de ORC y/o de los complejos MCM en regiones de cromatina más accesible.

En el punto anterior de la Discusión hemos comentado la posibilidad de que la secuencia de DNA impacte de manera directa (a través de estructuras de G4) o indirectamente (facilitando la aparición de *R-loops*) en el reconocimiento de ORC. Es posible, además, que la estructura del DNA generada por estas secuencias sea refractaria al empaquetamiento en nucleosomas que, a su vez, podría facilitar la

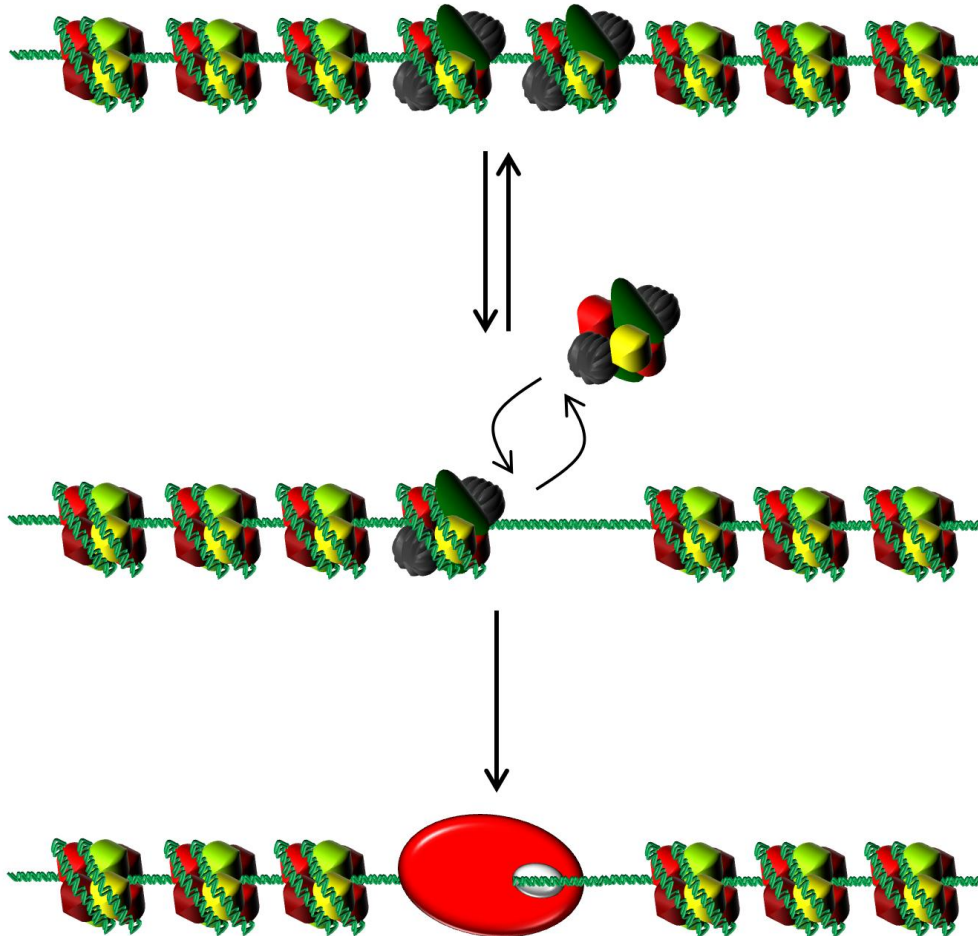


Figura 35: Modelo de la unión de ORC a regiones con nucleosomas lábiles. La presencia de variantes de histonas como H2A.Z y H3.3 en los nucleosomas (centrales) les confiere una inestabilidad que provoca un constante recambio en contraposición con los nucleosomas canónicos (laterales). Esto permite que factores de transcripción puedan acceder a los promotores (Jin *et al.*, 2009) y, probablemente, el mismo efecto es aprovechado por complejos proteicos como ORC (rojo) para unirse al DNA.

accesibilidad de los pre-RC a estas regiones. En la idea de que la secuencia puede determinar ciertas características de los procesos nucleares así como de la estructura de la cromatina, estudios recientes en tres especies de *Schizosaccharomyces* y en *S. cerevisiae* han descrito una huella en la secuencia del DNA que podría modular la incorporación de nucleosomas mediante una posición constante de bases A/T que además presentarían una asimetría en la cadena (*AT-skew*) (Quintales *et al.*, 2015). Debido a que el contenido A/T a lo largo de los genes es menor que en regiones intergénicas, esta huella aparece más asociada a estas últimas. Avanzando en esta propuesta, el empaquetamiento *in vivo* de DNA sintético conteniendo estas huellas en la secuencia reproducía los resultados postulados en el anterior trabajo de este

laboratorio (González *et al.*, 2016). Curiosamente en, *L. major*, los ORI más eficientes se localizan en regiones que contienen una base modificada, la base J, y una baja ocupación nucleosomal (**Figuras 24, 31 y 32**). De hecho, la observación individualizada cada una de estas regiones mediante mapas de color (**Figura 32**), confirma que la menor presencia de nucleosomas reproduce el patrón marcado por los máximos de detección de la base J y a su vez, esta coincidencia espacial, flanquea los mayores enriquecimientos en SNS. Podemos observar también este efecto en los ejemplos representativos de la **Figura 26**. Aunque los experimentos destinados a identificar la localización genómica de la proteína Orc1/Cdc6 en *L. major* han sido infructuosos hasta el momento, es posible que la cromatina diferencial de las regiones que contienen la base J favorezca el reclutamiento de ORC y, por tanto, la iniciación de la replicación en esas regiones. Tampoco sería descabellado sugerir que la base J participase de algún modo facilitando este reconocimiento ya que en *Trypanosoma cruzi* se ha detectado enriquecimiento de Orc1/Cdc6 en los telómeros, en los cuales se detecta el 99% de la base J en los promastigotes de *L. major* (van Luenen *et al.*, 2012; Tiengwe *et al.*, 2014).

3. La actividad transcripcional condiciona el perfil replicativo

Trabajos anteriores del laboratorio habían descrito la estrecha relación entre las regiones de inicio de la replicación y de la transcripción, demostrando que los ORI más eficientes en células ES de ratón se encontraban asociados a promotores (Sequeira-Mendes *et al.*, 2009). De acuerdo con estas observaciones, en los ORI cuya actividad hemos analizado a alta resolución también detectamos una correlación positiva entre la eficiencia de iniciación y la asociación a regiones promotoras (**Figura 8**). Como ya hemos comentado, una explicación plausible es la unión oportunista de los pre-RC a través de una cromatina accesible a los factores de transcripción y a la RNAPII. Esta interpretación concuerda con resultados de estudios genómicos recientes en células humanas en los que se han asociado 8.000 sitios de unión de ORC1 con promotores y *enhancers* (Dellino *et al.*, 2013) o hasta 50.000 regiones reconocidas por ORC2 que contienen alta concentración de factores de transcripción y están enriquecidos en marcas de histonas relacionadas con la iniciación de la transcripción, como la metilación de la lisina H3K4 (Miotto *et al.*, 2016). Curiosamente, en *L. major* esta relación entre ambos procesos no ocurre en las regiones de inicio de la transcripción de la RNAPII, sino que los ORI más eficientes se concentran en las zonas de terminación, en las que se también se localiza la base J, que contribuye al

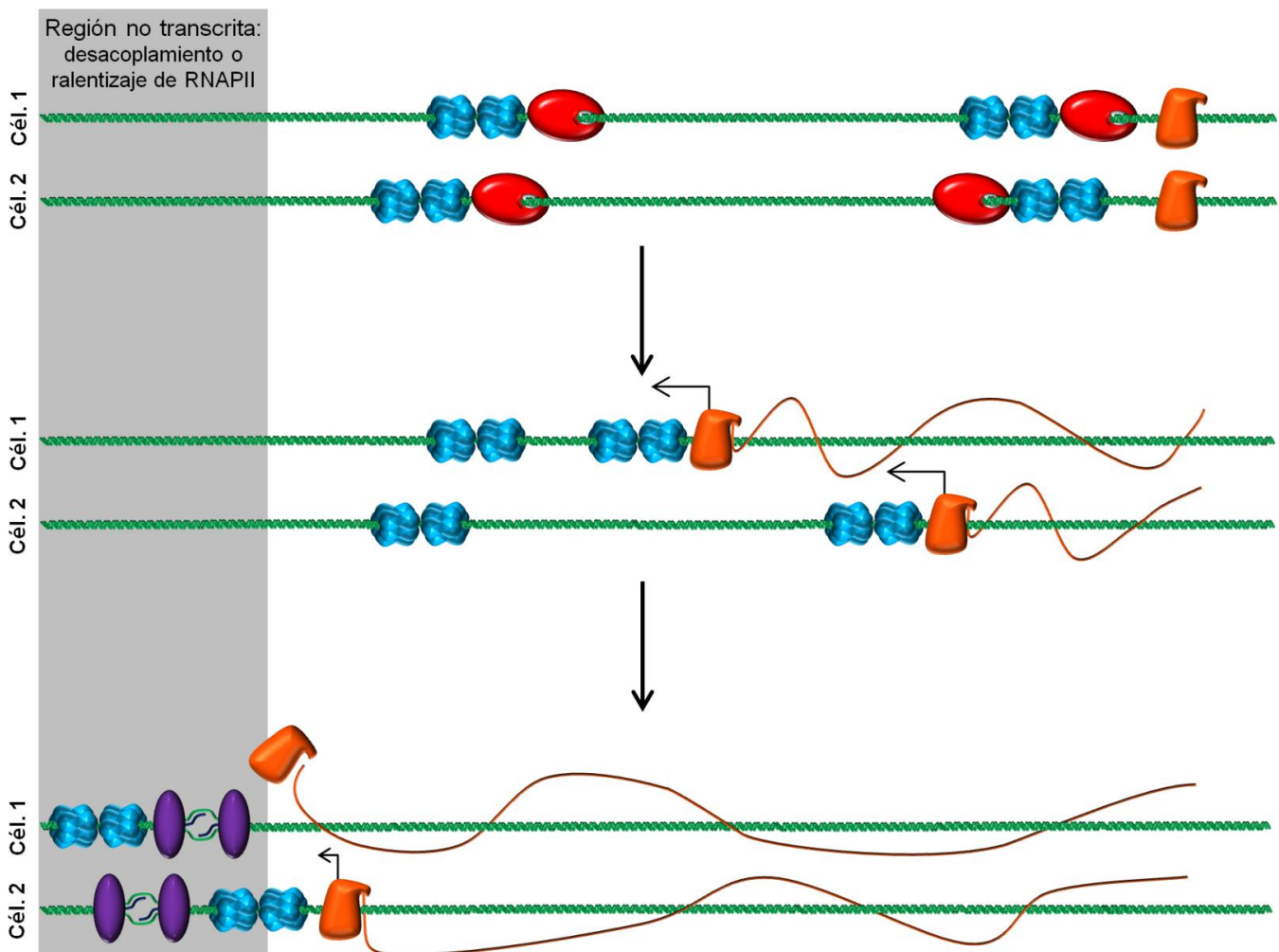
desensamblaje de la RNAPII impidiendo que continúe la transcripción de una PTU sobre la adyacente (van Luenen *et al.*, 2012) (**Figura 31A,C**).

En casos de sistemas en los que la transcripción está silenciada, como embriones tempranos de *D. melanogaster* o de *X. laevis*, la replicación se inicia de manera aleatoria a lo largo de todo el genoma (Shinomiya e Ina, 1991; Hyrien y Méchali, 1993). Sin embargo, en estados más avanzados del desarrollo en los que comienza a producirse transcripción génica, el inicio de la replicación se ve desplazado hacia regiones intergénicas (Hyrien *et al.*, 1995; Sasaki *et al.*, 1999). Esto sugiere que el tránsito de la RNAPII a lo largo del genoma podría afectar a la actividad de los ORI. Dado que la replicación del DNA ocurre una única vez por ciclo celular mientras que la maquinaria de transcripción recorre la doble hélice en múltiples ocasiones, es razonable pensar que el primer proceso sea modulado por el segundo. *L. major*, donde la transcripción por la RNAPII se realiza de manera constante en más del 90% del genoma, podría ser un modelo útil para profundizar en el estudio de este mecanismo.

Aunque la unión del complejo ORC al DNA es el primer paso en la cascada de licenciamiento y su localización genómica se utiliza como indicación de la presencia de un posible ORI (MacAlpine *et al.*, 2010; Dellino *et al.*, 2013), estudios recientes han demostrado que ORC es prescindible una vez que se han incorporado los complejos MCM a la cromatina. De hecho, en experimentos de reconstitución *in vitro* de los pre-RC de *S. cerevisiae* ha sido posible extraer ORC del DNA -incrementando la concentración de sal en condiciones en las que los complejos MCM aun permanecen unidos- manteniendo intacta su capacidad para iniciar la síntesis del DNA (Gros *et al.*, 2014). El reclutamiento de los dos hexámeros MCM al DNA ocurre adyacentemente al sitio de unión de ORC (Evrin *et al.*, 2009; Remus *et al.*, 2009). Esto explica que el inicio de la replicación detectado mediante SNS flanquee la posición de ORC (**Figura 14**). De manera destacable, estudios recientes en *S. cerevisiae* han demostrado que en células deficientes en terminación de la transcripción, colisiones con la RNAPII inducen la redistribución de los complejos MCM en el genoma, dando como resultado un desplazamiento de los sitios de iniciación de la replicación (Gros *et al.*, 2015). Paralelamente, un estudio en células Kc de *D. melanogaster* bloqueadas en la transición G1/S ha mostrado que los complejos MCM se redistribuyen desde los sitios de unión de ORC hacia regiones no transcritas del genoma (Powell *et al.*, 2015). Esta novedosa separación espacial entre las fases de reconocimiento del ORI de su licenciamiento, podría explicar el patrón bimodal de abundancia de SNS en células de mamífero debido al efecto de la elongación de la RNAPII a ambos lados del TSS

(Figuras 7 y 13). Basándonos en estos antecedentes postulamos que en *L. major* pudiera ocurrir un proceso similar: la actividad de la RNAPII podría especificar la iniciación de la replicación mediante el desplazamiento de los complejos MCM a lo largo de las PTU (Figura 36). Este escenario, aunque altamente especulativo, explicaría la menor presencia de ORI eficientes a lo largo de las PTU y su acumulación en las regiones de terminación de la transcripción (Figuras 24, 31 y 32). De hecho, en el análisis de regiones individuales mediante mapas de color (Figura 32), la coincidencia espacial entre máximos de localización de la base J y de abundancia de SNS, así como de mínima ocupación nucleosomal, ocurre en posición 3' de los TTS. Una explicación alternativa sería que, simplemente, la replicación en este organismo se inicia de manera oportunista en aquellos sitios que el proceso de transcripción ha dejado accesibles. Así, el programa replicativo de *L. major* emergería como una respuesta adaptativa a las restricciones en la cromatina impuestas por la transcripción. Sin duda, una interesante manera de investigar la importancia de las interacciones entre el inicio de la replicación del DNA y la terminación de la transcripción sería la deslocalización de ésta última, por ejemplo, eliminando la base J de los TTS. En este escenario se esperaría la deslocalización de la iniciación de la replicación en los ORI que actualmente percibimos como más eficientes o incluso, la mayor activación de aquellos cuyo conflicto con la transcripción resulte menos inconveniente en este nuevo panorama transcripcional. Además, el análisis detallado de la localización de los complejos Orc1/Cdc6 y MCM mediante CHIP en ambos escenarios añadiría una información esencial al conocimiento de cómo las células resuelven las interferencias entre estos procesos.

El hipotético desplazamiento de los complejos MCM hacia los TTS mediada por la elongación de la RNAPII también podría explicar la localización de los ORI que aparecen intercalados a lo largo de las PTU y que ocurre mayoritariamente en las regiones genómicas correspondientes a los sitios de *trans-splicing* entre los pre-RNAm (Figura 31B). La colisión de los complejos de elongación con barreras genómicas naturales, como los nucleosomas posicionados en estas zonas o la presencia de estructuras de G4, podría tener como consecuencia la pausa y, en algunas células, el eventual desensamblaje de la RNAPII, que dejaría sobre los cromosomas en esas posiciones a los complejos MCM (Figura 36). En diversos sistemas, se ha demostrado que el *splicing* se produce de manera co-transcripcional (revisado en Perales y Bentley, 2009). Por ejemplo, la RNAPII en *S. cerevisiae* se ralentiza en la frontera exón-intrón de manera que se facilita la interacción con los complejos de *splicing* y su resolución. Además, distorsiones tanto en la tasa de transcripción, como en el



ensamblaje de los complejos de *splicing*, provocan la ralentización de ambos procesos

Figura 36: Modelo explicativo de cómo la transcripción puede dirigir el inicio de la replicación en *L. major*. Tras el reconocimiento de los ORI por ORC (rojo), los complejos MCM (azul) se pueden mantener sobre el genoma aunque son desplazados por el paso de la RNAPII (naranja) (Gros *et al.*, 2014; 2015) acumulándose en zonas no transcritas (región sombreada) (Powell *et al.*, 2015). En regiones de fin de la transcripción (SSRc o HT) -célula 1-, la acumulación de MCM sobre ellas se produce en un alto porcentaje de los individuos de la población causando una frecuencia relativa de activación mayor -ORI eficientes-. Cuando la RNAPII modera su velocidad (*trans-splicing*, remodelación de la cromatina, etc.) -célula 2- los MCM desplazados y acumulados por la RNAPII pueden eventualmente ser activados en algunos individuos y, aunque esto ocurre en mayor número de ocasiones, la frecuencia relativa de activación de cada origen individual en la población es menor -ORI poco eficientes-.

(Carrillo-Oesterreich *et al.*, 2010; Carrillo-Oesterreich *et al.*, 2016). Esta interpretación acomodaría el hecho de que, aunque la mayoría de los ORI mapeados en *L. major* se encuentren en estas posiciones, la iniciación de la replicación ocurra en menos del 20% de todas las regiones de *trans-splicing* del genoma y que además, estos ORI

sean utilizados de manera menos eficiente en la población celular. En conjunto, nuestros resultados sugieren que los ORI en *L. major* no existirían como *loci* genéticos bien definidos. Esta conclusión es consistente con la observación de que este parásito es capaz de mantener la replicación de plásmidos que contengan cualquier porción de su genoma o de genomas exógenos (Patnaik *et al.*, 1997). Más bien parecería que en cada célula existe una cierta probabilidad de encontrar un ORI en una localización determinada, probabilidad que estaría determinada por la cinética de la transcripción de la RNAPII de cada célula y que a nivel poblacional es máxima en las regiones de TTS. De acuerdo con esta hipótesis, la probabilidad de encontrar un ORI en la región 3'UTR de un gen resultó ser mayor para los pre-RNAm más abundantes frente a los de estabilidad media (36% respecto a 25%, $p < 0.00001$, test χ^2), como se esperaría si la RNAPII se detuviera preferentemente en aquellas posiciones al encontrarse con un nucleosoma bien posicionado (**Figura 29D**). Esta correlación, sin embargo, no se daba al analizar el mismo fenómeno en el extremo 5'UTR de los genes ($p=0.2$).

Aunque los resultados en *L. major* no nos permiten establecer una relación causal entre la elongación de la transcripción y la actividad de los ORI, son consistentes con la interpretación de que el programa replicativo de este sistema podría explicarse como resultado de su asociación con la dinámica transcripcional de la RNAPII, sin necesidad de invocar mecanismos regulatorios específicos. Si esta hipótesis fuera correcta, esta solución podría aportar a este organismo la flexibilidad y robustez requerida para responder a los cambios ambientales que imponen las alteraciones en el programa genético debido a su ciclo de vida. Investigar si este escenario es una curiosidad de este parásito o si representa una solución ancestral utilizada por las células eucariotas para el mantenimiento de los ORI constituye una interesante propuesta para el futuro.

* * *

Al igual que otros miembros de la familia de los tripanosomátidos, *L. major* presenta un dimorfismo claramente definido a lo largo de su ciclo vital dividido en promastigote -estado libre, con movilidad flagelar, que presenta en los vectores de transmisión *Phlebotomus* y *Lutzomyia*- y amastigote -estado que adquiere al infectar al hospedador-. En nuestros estudios, hemos trabajado con el estado promastigote, mucho más manejable y sencillo de aislar que el patógeno amastigote. Estudios sobre

la estabilidad de los RNAm mediante *microarrays* determinan que las enormes diferencias fenotípicas entre ambos estados podrían deberse al cambio de estabilidad del 5-10% de los RNAm (Holtzer *et al.*, 2006). Aunque una gran parte de los cambios metabólicos serían dependientes de la regulación post-transcripcional (Holtzer *et al.*, 2006; McNicoll *et al.*, 2006), una posibilidad interesante sería analizar cómo de distinto es el paisaje replicativo de células amastigotes respecto a las promastigotes y, más aún, la dinámica de la estructura de la cromatina en aquellos pre-RNAm cuya estabilidad se ve modificada.

El estudio de la replicación en un modelo como *L. major* también podría representar un nuevo frente de actuación respecto a enfermedades infecciosas provocadas por estos parásitos. Además de las diferencias entre las maquinarias de replicación de estos respecto a su hospedador, estudiar en detalle cómo se transmite la información posicional de la base J a las cadenas recién sintetizadas del DNA podría descubrir nuevas dianas terapéuticas para combatir estas patologías.



CONCLUSIONES

Las conclusiones principales que podemos extraer de los trabajos que conforman esta Tesis Doctoral son:

1. Los orígenes de replicación más eficientes en células de mamífero contienen nucleosomas lábiles compuestos por dobles variantes de las histonas H2A.Z y H3.3 que facilitan el reclutamiento del complejo ORC a estas regiones. La síntesis del DNA comienza en las regiones inmediatamente adyacentes, en localizaciones que coinciden con nucleosomas bien posicionados.
2. Los ORI más eficientes contienen motivos en la secuencia del DNA con el potencial de formar estructuras secundarias como G4 o *R-loops*, asimetrías en la composición de bases, o la base modificada J en el caso del modelo promastigote de *Leishmania major*.
3. El paisaje nucleosomal en *L. major* refleja la actividad transcripcional de su genoma: bajo nivel de posicionamiento a lo largo de las PTU, exceptuando las regiones que corresponden a los sitios de *trans-splicing*.
4. El perfil replicativo en *L. major* está estrechamente relacionado con la cinética de la transcripción, detectándose la mayoría de los ORI en las regiones de ralentización o desensamblaje de la RNAPII, correspondiendo estos últimos a la categoría de ORI más eficientes.



BIBLIOGRAFÍA

- Abdurashidova G., Deganauto M., Klima R., Riva S., Biamonti G., Giacca M. y Falaschi A. (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin *Science* **287**: 2023-2026
- Agabian N. (1990) Trans-splicing of nuclear pre-mRNAs. *Cell* **61**: 1157-1160
- Agami R. y Shapira M. (1992) Nucleotide sequence of the spliced leader RNA gene from *Leishmania mexicana amazonensis*. *Nucl Acids Res* **20**: 1804
- Alabert C. y Groth A (2012) Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* **13**: 153-167
- Aladjem MI. (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet* **8**: 588-600
- Antequera F. (2003) Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* **60**: 1647-1658
- Antequera F. (2004) Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J* **23**: 4365-4370
- Aslett M., Aurrecochea C., Berriman M., Brestelli J., Brunk BP., Carrington M., Depledge DP., Fischer S., Gajria B., Gao X., *et al.* (2009) TriTrypDB: a functional genomic resource for the *Trypanosomatidae* *Nucl Acids Res* **38**(Database issue): 457-462
- Austin RJ., Orr-Weaver TL. y Bell SP. (1999) *Drosophila* ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* **13**: 2639-2649
- Baldauf SL. (2003) The deep roots of eukaryotes. *Science* **300**: 1703-1706
- Barry ER. y Bell SD (2006) DNA replication in archaea. *Microbiol Mol Biol Rev* **70**: 876-887
- Barski A., Cuddapah S., Cui K., Roh TY., Schones DE., Wang Z., Wei G., Chepelev I. y Zhao K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823-837
- Bell SP. y Stillman B. (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**: 128-134
- Berdis AJ. (2008) DNA polymerases as therapeutic targets. *Biochemistry* **47**: 8253-8260
- Besnard E., Babled A., Lapasset L., Milhavet O., Parrinello H., Dantec C., Marin JM. y Lemaitre JM. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* **19**: 837-844
- Bicknell LS., Bongers EM., Leitch A., Brown S., Schoots J., Harley ME., Aftimos S., Al-Aama JY., Bober M., Brown PA., *et al.* (2011) Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet* **43**: 356-359

- Bielinsky AK. (2003) Replication origins: why do we need so many? *Cell Cycle* **2**: 307-309
- Bird A., Taggart M., Frommer M., Miller OJ. y Macleod D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**: 91-99
- Blumenthal AB. Kriegstein HJ. y Hogness DS. (1974) The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harbor Symp Quant Biol* **38**: 205-223
- Boos D., Frigola J. y Diffley JF. (2012) Activation of the replicative DNA helicase: breaking up is hard to do. *Curr Opin Cell Biol* **24**: 423-430
- Bugaut A. y Balasubramanian S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **47**: 689-697
- Burge S., Parkinson GN., Hazel P. Todd AK. Y Neidle S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucl Acids Res* **34**: 5402-5415
- Cadoret JC., Meisch F., Hassan-Zadeh V., Luyten I., Guillet C., Duret L., Quesneville H. y Prioleau MN. (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci USA* **105**: 15837-15842
- Cahoon LA. y Seifert HS. (2013) Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in *Neisseria gonorrhoeae*. *PLoS Pathog* **9**: e1003074
- Carrillo-Oesterreich F., Herzel L., Straube K., Hujer K., Howard J. y Neugebauer KM. (2016) Splicing of nascent RNA coincides with intron exit from RNA Polymerase II. *Cell* **165**: 372-381
- Carrillo-Oesterreich F., Preibisch S. y Neugebauer KM. (2010) Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**: 571-581
- Cayrou C., Ballester B., Peiffer I., Fenouil R., Coulombe P., Andrau JC., van Helden J. y Méchali M. (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res* **25**: 1873-1885
- Cayrou C., Coulombe P., Puy A., Rialle S., Kaplan N., Segal E. y Méchali M. (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle* **11**: 658-667
- Cayrou C., Coulombe P., Vigneron A., Stanojcic S., Ganier O., Peiffer I., Rivals E., Puy A., Laurent-Chabalier S., Desprat R. y Méchali M. (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* **21**: 1438-1449
- Chambers VS., Marsico G., Boutell JM., di Antonio M., Smith GP. y Balasubramanian S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* **33**: 877-881

- Chang DD., Hauswirth WW. y Clayton DA. (1985) Replication priming and transcription initiate from precisely the same site in mouse mitochondrial DNA. *EMBO J* **4**:1559-1567
- Christov CP., Gardiner TJ., Szüts D. y Krude T. (2006) Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. *Mol Cell Biol* **26**: 6993-7004
- Chuang RY. y Kelly TJ. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci USA* **96**: 2656-2661
- Clemente-Ruiz M. y Prado F. (2009) Chromatin assembly controls replication fork stability. *EMBO Rep* **10**: 790-796
- Comoglio F., Schlumpf T., Schmid V., Rohs R., Beisel C. y Paro R. (2015) High-resolution profiling of *Drosophila* replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep* **11**: 821-834
- Costa A., Ilves I., Tamberg N., Petojevic T., Nogales E., Botchan MR. y Berger JM. (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* **18**: 471-477
- Costas C., de la Paz-Sánchez M., Stroud H., Yu Y., Oliveros JC., Feng S., Benguria A., López-Vidriero I., Zhang X., Solano R., *et al.* (2011) Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nat Struct Mol Biol* **18**: 395-400
- Creyghton MP., Cheng AW., Welstead GG., Kooistra T., Carey BW., Steine EJ., Hanna J., Lodato MA., Frampton GM., Sharp PA., *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**: 21931-21936
- Cuadrado M., Sacristán M. y Antequera F. (2001) Species-specific organization of CpG island promoters at mammalian homologous genes. *EMBO Rep* **2**: 586-592
- Dahmann C., Diffley JF. y Nasmyth KA. (1995) S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. *Curr Biol* **5**: 1257-1269
- de Castro E., Soriano I., Marín L., Serrano R., Quintales L. y Antequera F. (2012) Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. *EMBO J* **31**: 124-137
- de Munnik SA., Bicknell LS., Aftimos S., Al-Aama JY., van Bever Y., Bober MB., Clayton-Smith J., Edrees AY., Feingold M., Fryer A., *et al.* (2012) Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex mutations and 10 without molecular diagnosis. *Eur J Hum Genet* **20**: 598-606
- Deal RB. y Henikoff S. (2010) Capturing the dynamic epigenome. *Genome Biol* **11**: 218-225

- Delgado S., Gómez M., Bird A., y Antequera F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J* **17**: 2426-2435
- Dellino GI., Cittaro D., Piccioni R., Luzi L., Banfi S., Segalla S., Cesaroni M., Mendoza-Maldonado R., Giacca M. y Pelici PG. (2013) Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res* **23**: 1–11
- DePamphilis ML. (1996) Origins of DNA replication. En: DePamphilis ML, DNA replication in eukaryotic cells. *CSHL Press* 45-86
- DePamphilis ML. (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* **21**: 5-16
- DePamphilis ML. (2005) Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle* **4**: 70-79
- Diffley JF. (2004) Regulation of early events in chromosome replication. *Curr Biol* **14**: 778-786
- Diffley JF. (2011) Quality control in the initiation of eukaryotic DNA replication. *Philos Trans R Soc Lond B Biol Sci* **366**: 3545-3553
- Diffley JF. y Stillman B. (1988) Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. *Proc Natl Acad Sci USA* **85**: 2120-2124
- Dijkwel PA., Vaughn JP. y Hamlin JL. (2002) Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. *Mol Cell Biol* **22**: 3053-3065
- Dubey DD., Kim SM., Todorov IT. y Huberman JA. (1996) Large, complex modular structure of a fission yeast DNA replication origin. *Curr Biol* **6**: 467-473
- Eaton ML. Prinz JA., MacAlpine HK., Tretyakov G., Kharchenko PV. y MacAlpine DM. (2011) Chromatin signatures of the *Drosophila* replication program. *Genome Res* **21**: 164-174
- Eaton ML., Galani K., Kang S., Bell SP. y MacAlpine DM. (2010) Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**: 748-753
- Ehrlich M., Gama-Sosa MA., Huang LH., Midgett RM., Kuo KC., McCune RA. y Gehrke C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucl Acids Res* **10**: 2709-2721
- Evrin C., Clarke P., Zech J., Lurz R., Sun J., Uhle S., Li H., Stillman B. y Speck C. (2009) A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication *Proc Natl Acad Sci USA* **106**: 20240-20245
- Fan X., Moqtaderi Z., Jin Y., Zhang Y., Liu XS., y Struhl K. (2010) Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *Proc Natl Acad Sci USA* **107**: 17945-17950

- Farkas G., Leibovitch BA. y Elgin SC. (2000) Chromatin organization and transcriptional control of gene expression in *Drosophila*. *Gene* **253**: 117-136
- Feng W., Collingwood D., Boeck ME., Fox LA., Alvino GM., Fangman WL., Raghuraman MK. y Brewer BJ. (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeast identifies origins of replication. *Nat Cell Biol* **8**: 148-155
- Filion GJ., van Bommel JG., Braunschweig U., Talhout W., Kind J., Ward LD., Brugman W., de Castro IJ., Kerkhoven RM., Bussemaker HJ. y van Steensel B. (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**: 212-224
- Fouk MS., Urban JM., Casella C. y Gerbi SA. (2015) Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res* **25**: 725-735
- Fouse SD., Shen Y., Pellegrini M., Cole S., Meissner A., van Neste L., Jaenisch R. y Fan G. (2008) Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**: 160-169
- Fragkos M., Ganier O., Coulombe P. y Méchali M. (2015) DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* **16**: 360-374
- Gambus A., Jones RC., Sánchez-Díaz A., Kanemaki M., van Deursen F., Edmondson RD. y Labib K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* **8**: 358-366
- Gambus A., Khoudoli GA., Jones RC. y Blow JJ. (2011) MCM2-7 form double hexamers at licensed origins in *Xenopus* eggs extract. *J Biol Chem* **286**: 11855-11864
- Gambus A., van Deursen F., Polychronopoulos D., Foltman M., Jones RC., Edmondson RD., Calzada A. y Labib K. (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *EMBO J* **28**: 2992-3004
- Giacca M., Zentilin L., Norio P., Diviacco S., Dimitrova D., Contreas., Biamonti G., Perini G., Weighardt F. y Riva S. (1994) Fine mapping of a replication origin of human DNA. *Proc Natl Acad Sci USA* **91**: 7119-7123
- Giannoulatou E., Park SH., Humphreys DT. y Ho JW. (2014) Verification and validation of bioinformatics software without a gold standard: a case study of BWA and Bowtie. *BMC Bioinformatics* **15** Suppl 16: S15
- Gindin Y., Valenzuela MS., Aladjem MI., Meltzer PS. y Bilke S. (2014) A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Mol Syst Biol* **10**: 722
- Ginno PA., Lim YW., Lott PL., Korf I. y Chedin F. (2013) GC skew at the 5' and 3' ends of human cell links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* **23**: 1590-1600

- GINNO PA., LOTT PL., CHRISTENSEN HC., KORF I. y CHEDIN F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**: 814-825
- GODOY PD., NOGUEIRA-JUNIOR LA., PAES LS., CORNEJO A., MARTINS RM., SILBER AM., SCHENKMAN S. y ELIAS MC. (2009) Trypanosome prereplication machinery contains a single functional orc1/cdc6 protein, which is typical of archaea. *Eukaryot Cell* **8**: 1592-1603
- GOLDMAN MA., HOLMQUIST GP., GRAY MC., CASTON LA. y NAG A. (1984) Replication timing of genes and middle repetitive sequences. *Science* **224**: 686-692
- GÓMEZ M y ANTEQUERA F. (2008) Overreplication of short DNA regions during S phase in human cells. *Genes Dev* **22**: 375-385
- GOMEZ M. y ANTEQUERA F. (1999) Organization of DNA replication origins in the fission yeast genome. *EMBO J* **18**: 5683-5690
- GOMMERS-AMPT JH., VAN LEEUWEN F., DE BEER AL., Vliegenthart JF., Dizdaroglu M., Kowalak JA., Crain PF. y Borst P. (1993) β -D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* **75**: 1129-1136
- GONG QH., MCDOWELL JC. y DEAN A. (1996) Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene *in vivo* by 5' hypersensitive site 2 of the beta-globin locus control region. *Mol Cell Biol* **16**: 6055-6064
- GONZÁLEZ S., GARCÍA A., VÁZQUEZ E., SERRANO R., SÁNCHEZ M., QUINTALES L. y ANTEQUERA F. (2016) Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome. *Genome Res* **26**: 1532-1543
- GROS J., DEVBHANDARI S. y REMUS D. (2014) Origin plasticity during budding yeast DNA replication *in vitro*. *EMBO J* **33**: 621-636
- GROS J., KUMAR C., LYNCH G., YADAV T., WHITEHOUSE I. y REMUS D. (2015) Post-licensing specification of eukaryotic replication origins by facilitated Mcm2-7 sliding along DNA. *Mol Cell* **60**: 797-807
- GUÉDIN A., GROS J., ALBERTI P. y MERGNY JL. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucl Acids Res* **38**: 7858-7868
- GUSCHLBAUER W., CHANTOT JF. y THEILE D. (1990) Four-stranded nucleic structures 25 years later: from guanosine gels to telomere DNA. *J Biomol Struct Dyn* **8**: 8491-8511
- HAZEL P., HUPPERT J., BALASUBRAMANIAN S. y NEIDLE S. (2004) Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc* **126**: 16405-16415
- HEINTZ NH. y HAMLIN JL. An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific restriction fragments. *Proc Natl Acad Sci USA* **79**: 4083-4087

- Heinzel SS., Krysan PJ., Tran CT. y Calos MP. (1991) Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Mol Cell Biol* **11**: 2263-2272
- Henikoff JG., Belsky JA., Krassovsky K., MacAlpine DM. y Henikoff S. (2011) Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci USA* **108**: 18318-18323
- Holzer TR., McMaster WR. y Forney JD. (2006) Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes and axenic amastigotes in *Leishmania mexicana*. *Mol Biochem Parasitol* **146**: 198-218
- Hoshina S., Yura K., Teranishi H., Kiyasu N., Tominaga A., Kadoma H., Nakatsuka A., Kunichika T., Obuse C. y Waga S. (2013) Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J Biol Chem* **288**: 30161-10171
- Huberman JA. y Riggs AD. (1968) On the mechanism of DNA replication in mammalian chromosomes. *J Mol Biol* **32**: 327-341
- Huppert JL. y Balasubramanian S. (2005) Prevalence of quadruplexes in the human genome. *Nucl Acids Res* **33**: 2908–2916
- Hyrien O. y Méchali M. (1993) Chromosomal replication initiates and terminates at random sequences but at regular intervals intervals in the ribosomal DNA of *Xenopus* early embryos. *EMBO J* **12**: 4511-4520
- Hyrien O., Maric C. y Méchali M. (1995) Transition in specification of embryonic metazoan DNA replication origins. *Science* **270**: 994-997
- Ibarra A., Schwob E. y Méndez J. (2008) Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci USA* **105**: 8956-8961
- Illingworth RS. y Bird AP. (2009) CpG islands: a rough guide *FEBS Lett* **583**: 1713-1720
- Illingworth RS., Gruenewald-Schneider U., Webb S., Kerr AR., James KD., Turner DJ., Smith C., Harrison DJ., Andrews R. y Bird A. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**: e1001134
- Ilves I., Petojevic T., Pesavento JJ. y Botchan MR. (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* **37**: 247-258
- Ivens AC., Peacock CS., Worthley EA., Murphy L., Aggarwal G., Berriman M., Sisk E., Rajandream MA., Adlem E., Aert R. *et al.* (2005) The genome of the kinetoplastid parasite *Leishmania major*. *Science* **309**: 436-442
- Jacob F., Brenner S. y Cuzin F. (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp Quant Biol* **28**: 329-348

- Jenuwein T. y Allis CD. (2001) Translating the histone code. *Science* **293**: 1074-1080
- Jiang C. y Pugh BF. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**: 161-172
- Jin C. y Felsenfeld G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* **21**: 1519–1529
- Jin C., Zang C., Wei G., Cui K., Peng W., Zhao K. y Felsenfeld G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nat Genet* **41**: 941–945
- Kaykov A., Taillefumier T., Bensimon A. y Nurse P. (2016) Molecular combing of single DNA molecules on the 10 megabase scale. *Sci Rep* **19**: 6:19636
- Keller C., Ladenburger EM., Kremer M. y Knippers R. (2002) The origin recognition complex marks a replication origin in the human TOP1 gene promoter. *J Biol Chem* **277**: 31430-31440
- Kelman LM. y Kelman Z. (2014) Archaeal DNA replication. *Annu Rev Genet* **48**: 71-97
- Kent WJ., Sugnet CW., Furey TS., Rosking KM., Pringle TH., Zahler AM. y Haussler D. (2002) The human genome browser at UCSC. *Genome Res* **12**: 996-1006
- Kharchenko PV., Alekseyenko AA., Schwartz YB., Minoda A., Riddle NC., Ernst J., Sabo PJ., Larschan E., Gorchakov AA., Gu T., *et al.* (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480-485
- Kitsberg d., Selig S., Keshet I. y Cedar H. (1993) Replication structure of the human β -globin gene domain. *Nature* **366**: 588-590
- Koch CM., Andrews RM., Flicek P., Dillon SC., Karaöz U., Clelland GK., Wilcox S., Beare DM., Fowler JC., Couttet P., *et al.* (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* **17**: 691-707
- Koren A., Handsaker RE., Kamitaki N., Karlic R., Ghosh S., Polak P., Eggan K. y McCarroll SA. (2014) Genetic variation in human DNA replication timing. *Cell* **159**: 1015-1026
- Kornberg RD. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* **184**: 868-871
- Kornberg RD. y Lorch Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**: 285-294
- Labib K., Tercero JA. y Diffley JF. (2000) Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science* **288**: 1643-1647
- Ladenburger EM., Keller C. y 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

- Langmead B. y Salzberg S. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359
- Langmead B., Trapnell C., Pop M. y Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25
- LeBowitz HG., Smith HQ., Rusche L. y Beverley SM. (1993) Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* **7**: 996-1007
- Lee W., Tillo D., Bray N., Morse RH., Davis RW., Hughes TR. y Nislow C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**:1235-1244
- Leonard AC. y Grimwade JE. (2015) The orisome: structure and function. *Front Microbiol* **6**: 545
- Li H. y Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. y 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079
- Li M., Hada A., Sen P., Olufemi L., Hall MA., Smith BY., Forth S., McKnight JN., Patel A., Bowman GD., *et al.* (2015) Dynamic regulation of transcription factors by nucleosome remodeling. *Elife* **4**
- Lipford JR. y Bell SP. (2001) Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol Cell* **7**: 21-30
- Logan-Klumpler F., de Silva N., Boehme U., Rogers MB., Velarde G., McQuillan JA., Carver T., Aslett M., Olsen C., Subramanian S., *et al.* (2012) GeneDB – an annotation database for pathogens. *Nucl Acids Res* **40**(Database issue): 98-108
- Lombraña R., Almeida R., Álvarez A. y Gómez M. (2015) R-loops and initiation of DNA replication in human cells: a missing link? *Front Genet* **6**: 158
- Lombraña R., Almeida R., Revuelta I., Madeira A., Herranz G., Saiz N., Bastolla U. y Gómez M. (2013) High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J* **32**: 2631-2644
- Lutzmann M. y Méchali M. (2009) How to load a replicative helicase onto chromatin: a more and more complex matter during evolution. *Cell Cycle* **8**: 1309-1313
- MacAlpine DM., Rodríguez HK. y Bell SP. (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev* **18**: 3094-3105
- MacAlpine HK., Gordân R., Powell SK., Hartemink AJ. y MacAlpine DM. (2010) *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* **20**: 201-211

- Maehara K. y Ohkawa Y. (2016) Exploration of nucleosome positioning patterns in transcription factor function. *Sci Rep* **6**: 19620
- Marahrens Y. y Stillman B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**: 817-823
- Marks AB., Smith OK. y Aladjem MI. (2016) Replication origins: determinants or consequences of nuclear organization? *Curr Opin Genet Dev* **37**: 67-75
- Marques CA., Dickens NJ., Paape D., Campbell SJ. y McCulloch R. (2015) Genome-wide mapping reveals single-origin chromosome replication in *Leishmania*, an eukaryotic microbe. *Genome Res* **20**: 201-211
- Martinez-Calvillo S., Nguyen D., Stuart K. y Myler PJ. (2004) Transcription initiation and termination on *Leishmania major* chromosome 3. *Eukaryot Cell* **3**: 506-517
- Martinez-Calvillo S., Yan S., Nguyen D., Fox M., Stuart K. y Myler PJ. (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* **11**: 1291-1299
- Masai H., Matsumoto S., You Z., Yoshizawa-Sugata N. y Oda M. (2010) Eukaryotic chromosome DNA replication: where, when and, how? *Annu Rev Biochem* **79**: 89-130
- Masukata H. y Tomizawa J. (1990) A mechanism of formation of a persistent hybrid between elongating RNA and template DNA *Cell* **62**: 331-338
- McNicoll F., Drummelstmith J., Müller M., Madore E., Boilard N., Ouellette M. y Papadopoulou B. (2006) A combined proteomic and transcriptomic approach to the study of stage differentiation in *Leishmania infantum*. *Proteomics* **6**: 3567-81
- Méchali M. (2010) Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* **11**: 728-738
- Méchali M. y Kearsley S. (1984) Lack of specific sequence requirement for DNA replication in *Xenopus* eggs compared with high sequence specificity in yeast. *Cell* **38**: 55-64
- Méndez J. y Stillman B. (2003) Perpetuating the double hélix: molecular machines at eukaryotic DNA replication origins. *Bioessays* **25**: 1158-1167
- Miotto B., Ji Z. y Struhl K. (2016) Selectivity of ORC binding sites and the relation to replication timing, fragile sites and deletions in cancers. *Proc Natl Acad Sci USA* **113**: 4810-4819
- Mojardin L., Vázquez E. y Antequera F. (2013) Specification of DNA replication origins and genomic base composition in fission yeasts. *J Mol Biol* **425**: 4706-4713
- Müller CA., Hawkins M., Retkute R., Malla S., Wilson R., Blythe MJ., Nakato R., Komata M., Shirahige K., de Moura APS. y Nieduszynski CA. (2014) The dynamics of genome replication using deep sequencing. *Nucl Acid Res* **42**: e3

- Mundade R., Ozer HG., Wei H., Prabhu L. y Lu T. (2014) Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* **13**: 2847-2852
- Muñoz S. y Méndez J. (2016) DNA replication stress: from molecular mechanisms to human disease. *Chromosoma* 1-15
- Musiak MW. y Rybaczek D. (2015) Behavior of replication origins in Eukaryota - spatio-temporal dynamics of licensing and firing. *Cell Cycle* **14**: 2251-2264
- Nakamura H. y Morita T. y Sato C. (1986) Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* **165**: 291-297
- Newlon CS. y Theis JF. (1993) The structure and function of yeast ARS elements. *Curr Opin Genet Dev* **3**: 752-758
- Nieduszynski CA., Knox Y. y Donaldson AD. (2006) Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* **20**: 1874-1879
- Norseen J., Johnson FB. y Lieberman PM. (2009) Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. *J Virol* **83**: 10336-10346: 10336-10346
- Norseen J., Thomae A., Sridharan V., Aiyar A., Schepers A. y Lieberman PM. (2008) RNA-dependent recruitment of the origin recognition complex. *EMBO J* **27**: 3024-3035
- Ozsolak F., Song JS., Liu XS. y Fisher DE. (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat Biotechnol* **25**: 244-248
- Pacek M., Tutter AV., Kubota Y., Takisawa H. y Walter JC. (2006) Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* **21**: 581-587
- Park PJ. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**: 669-680
- Patnaik PK. (1997) Studies with artificial extrachromosomal elements in trypanosomatids: could specificity in the initiation of DNA replication be linked to that in transcription? *Parasitol Today* **13**: 468-471
- Patrick E., Buckley M. y Yang YH. (2013) Estimation of data-specific constitutive exons with RNA-seq data. *BMC Bioinformatics* **14**: 31
- Penny GD., Kay GF., Sheardown SA., Rastan S. y Brockdorff N. (1996) Requirement for Xist in X chromosome inactivation. *Nature* **379**: 131-7
- Perales R. y Bentley D. (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**: 178-191
- Picard F., Cadoret JC., Audit B., Arneodo A., Alberti A., Battail C., Duret L. y Prioleau MN. (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet* **10**: e1004282

- Powell SK., MacAlpine HK., Printz JA., Li Y., Belsky JA. y MacAlpine DM. (2015) Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* **34**: 531-543
- Prioleau MN. y MacAlpine DM. (2016) DNA replication origins – Where do we begin? *Genes Dev* **30**: 1683-1697
- Prioleau MN., Gendron MC. y Hirien O. (2003) Replication of the chicken β -globin locus: early firing origins at the 5' HS4 insulator and the ρ - and β^A -globin genes show opposite epigenetic modifications. *Mol Cell Biol* **23**: 3536-3549
- Pruitt KD., Tatusova T. y Maglott DR. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucl Acids Res* **33**(Database issue): 501-504
- Quinlan AR y Hall IM. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842
- Quintales L., Vázquez E. y Antequera F. (2015) Comparative analysis of methods for genome-wide nucleosome cartography. *Brief Bioinf* **16**: 576-587
- R Core Team (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <http://www.Rproject.org>
- Raghuraman MK., Winzeler EA., Collingwood D., Hunt S., Wodicka L., Conway A., Lockhart DJ., Davis RW., Brewer BJ. y Fangman WL. (2001) Replication dynamics of the yeast genome. *Science* **294**: 115-121
- Randell JC., Fan A., Chan C., Francis LI., Heller RC., Galani K. y Bell SP. (2010) Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell* **40**: 353-363
- Rao H., Marahrens Y. y Stillman B. (1994) Functional conservation of multiple elements in yeast chromosomal replicators. *Mol Cell Biol* **14**: 7643-7651
- Rastrojo A., Carrasco-Ramiro F., Martín D., Crespillo A., Reguera RM., Aguado B. y Requena JM. (2013) The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq. *BMC Genomics* **14**: 223
- Remus D., Beuron F., Tolun G., Griffith JD., Morris EP. y Diffley JF. (2009) Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* **139**: 719-730
- Requena JM. (2011) Lights and shadows on gene organization and regulation of gene expression in *Leishmania*. *Front Biosci* **16**: 2069-2085
- Roberts RW. y Crothers DM (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition *Science* **258**: 1463-1466

- Robinson A., Causer RJ. y Dixon NE (2012) Architecture and conservation of the bacterial DNA replication, an underexploited drug target. *Curr Drug Targets* **13**: 352-372
- Robinson JT, Thorvaldsdóttir H, Winckler W., Guttman M., Lander ES., Getz G. y Merisov JP. (2011) Integrative Genomics Viewer. *Nature Biotechnology* **29**: 24-26
- Roudier F., Ahmed I., Bérard C., Sarazin A., Mary-Huard T., Cortijo S., Bouyer D., Caillieux E., Duvernois-Berthet E., Al-Shikhley L., *et al.* (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* **30**: 1928-1938
- Rubio-Camarillo M., Gómez-López G., Fernández JM., Valencia A. y Pisano DG. (2013) RUBioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses. *Bioinformatics* **29**: 1687-1689
- Sambrook J., Fritsch EF. y Maniatis T. (1989) Molecular Cloning – A Laboratory Manual. 2ª Ed. **CSHL Press**
- Sanz LA., Hartono SR., Lim YW., Steyaert S., Rajpurkar A., Ginno PA., Xu X. y Chédin F. (2016) Prevalent, dynamic and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell* **63**: 167-178
- Sasaki T., Sawado T., Yamaguchi M. y Shinomiya T. (1999) Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNAPol α -dE2F locus of *Drosophila melanogaster*. *Mol Cell Biol* **19**: 547-555
- Schones DE., Cui K., Cuddapah S., Roh TY., Barski A., Wang Z., Wei G. y Zhao K. (2008) Dynamic regulation of nucleosome positioning in the human genome *Cell* **132**: 887-898
- Segurado M., de Luis A. y Antequera F. (2003) Genome-wide distribution of DNA replication origins at AT-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* **4**: 1048-1053
- Sekinger EA., Moqtaderi Z. y Struhl K. (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* **18**: 735-748
- Sequeira-Mendes J. y Gómez M. (2012) On the opportunistic nature of transcription and replication initiation in the metazoan genome. *BioEssays* **34**: 119-125
- Sequeira-Mendes J., Aragüez I., Peiró R., Méndez Giráldez R., Zhang X., Jacobsen SE., Bastolla U. y Gutiérrez C. (2014) The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **26**: 2351-2366
- Sequeira-Mendes J., Díaz-Uriarte R., Apedaile A., Huntley D., Brockdorff N. and Gómez M. (2009) Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* **5**: e1000446
- Sharma S., Kelly TK. y Jones PA. (2010) Epigenetics in cáncer. *Carcinogenesis* **31**: 27-36

- Sheu YJ. y Stillman B. (2010) The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature* **463**: 113-117
- Shinomiya T. e Ina S. (1991) Analysis of chromosomal replicons in early embryos of *Drosophila melanogaster* by two-dimensional gel electrophoresis. *Nucl Acids Res* **19**: 3935-3941
- Smith DJ. y Whitehouse I. (2012) Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**: 434-438
- Spies N., Nielsen CB., Padgett RA. y Burge CB. (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36**: 245-254
- Stanojcic S., Lemaitre JM., Brodolin K., Danis E. y Méchali M. (2008) In *Xenopus* egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences. *Mol Cell Biol* **28**: 5265-5274
- Stiles JK., Hicock PI., Shah PH. y Meade JC. (1999) Genomic organization, transcription, splicing and gene regulation in *Leishmania*. *Ann Trop Med Parasitol* **93**: 781-807
- Stinchcomb DT. Struhl K. y Davis RW. (1979) Isolation and characterization of a yeast chromosomal replicator. *Nature* **282**: 39-43
- Stroud H., Otero S., Desvoyes B., Ramírez-Parra E., Jacobsen SE. y Gutiérrez C. (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **109**: 5370-5375
- Taira T., Iguchi-Ariga SM. y Ariga H. (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol Cell Biol* **14**: 6386-6397
- Tazi J. y Bird A. (1990) Alternative chromatin structure at CpG islands. *Cell* **60**: 909-920
- Teif VB., Vainshtein Y., Caudron-Herger M., Mallm JP., Marth C., Höfer T. y Rippe K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat Struct Mol Biol* **19**: 1185-1192
- Thomas S., Green A., Sturm NR. Campbell DA. y Myler PJ. (2009) histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* **10**: 152
- Tiengwe C., Marcello L., Farr H., Dickens N., Kelly S., Swiderski M., Vaughan D., Gull K., Barry JD., Bell SD. y McCulloch R. (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell rep* **26**: 185-197
- Tiengwe C., Marques CA. y McCulloch R. (2014) Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process. *Trends Parasitol* **30**: 27-36
- Tolstorukov MY., Kharchenko PV. y Park PJ. (2010) Analysis of primary structure of chromatin with next-generation sequencing. *Epigenomics* **2**: 187-197

- Valouev A., Johnson SM., Boyd SD., Smith CL., Fire AZ. y Sido A. (2011) Determinants of nucleosome organization in primary human cells. *Nature* **474**: 516-520
- Valton AL. y Prioleau MN. (2016) G-quadruplexes in DNA replication: a problem or a necessity? *Trends Genet* **32**: 697-706
- Valton AL., Hassan-Zadeh V., Lema I., Boggetto N., Alberti P., Saintomé C., Riou JF. y Prioleau MN (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J* **33**: 732-746
- Van Luenen HG., Farris C., Jan S., Genest PA., Tripathi P., Velds A., Kerkhoven RM., Nieuwland M., Haydock A. y Ramasamy G. (2012) Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in *Leishmania*. *Cell* **150**: 909-921
- Wigley DB. (2009) ORC proteins: marking the start. *Curr Opin Struct Biol* **19**: 72-78
- Woodward AM., Göhler T., Luciani MG., Oehlmann M., Ge X., Gartner A., Jackson DA. y Blow JJ. (2006) Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol* **173**: 673-683
- Wyrick JJ., Aparicio JG., Chen T., Barnett JD., Jennings EG., Young RA., Bell SP. y Aparicio OM. (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* **294**: 2357-2360
- Xu H., Handoko L., Wei X., Ye C., Sheng J., Wei CL., Lin F. y Sung WK. (2010) A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics* **26**: 1199-1204
- Yeeles JT., Deegan TD., Janska A. Early A. y Diffley JF. (2015) Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* **519**: 431-435
- Zakrzewska-Czerwinska J., Jakimowicz D., Zawilak-Pawlik A. y Messer W. (2007) Regulation of the initiation of chromosomal replication in bacteria. *FEMS Microbiol Rev* **31**: 378-387
- Zegerman P. y Diffley JF. (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* **445**: 281-285
- Zhang Y., Liu T., Meyer CA., Eeckhoute J., Johnson DS., Bernstein BE., Nusbaum C., Myers RM., Brown M., Li W. y Liu XS. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137

RATÓN				
NOMBRE	REGIÓN	CARACTERÍSTICAS	SECUENCIA (5' -> 3')	Tª
McpFl_5_Fw	Mecp2	Flanco 5'	TTGTGGCGCACTCTCCCAAC	66°C
McpFl_5_Rv			CACACAGACTGGCGCGCGTG	
Mcp1_Fw	Mecp2	Prom-ORI	GAGGGCGTCATTGGAAGGTG	64°C
Mcp1_Rv			AATCGCGAGCGACGGTTCTC	
Mcp2_Fw	Mecp2	Prom-ORI	GCTGAGAACCCTCGCTCGC	66°C
Mcp2_Rv			CGCCGGTGGTGGCTTTCTC	
Mcp3_Fw	Mecp2	Prom-ORI	GAGAAAGCCACCACCGGCG	66°C
Mcp3_Rv			CGACACGGCTGGCGGATG	
Mcp4_Fw	Mecp2	Prom-ORI	CATCCGCCAGCCGTGTGCG	60°C
Mcp4_Rv			CTTGCTGGGGGGCGGGTAG	
Mcp5_Fw	Mecp2	Prom-ORI	CTACCCGCCCCAGCAAG	64°C
Mcp5_Rv			GTGAGTGGGACCGCCAAGG	
Mcp6_Fw	Mecp2	Prom-ORI	CCTTGGCGGTCCCACTCAC	66°C
Mcp6_Rv			GCCGAGCGGAGGAGGAGG	
Mcp7_Fw	Mecp2	Prom-ORI	CTCCTCCGCTCGGCGCG	60°C
Mcp7_Rv			GCTGTGGTAAAACCCGTCCGG	
Mcp8_Fw	Mecp2	Prom-ORI	TTTTCCGGACGGTTTTACC	64°C
Mcp8_Rv			CGCTCCCTCCTCTCGGAG	
Mcp9_Fw	Mecp2	Prom-ORI	CTCCGAGAGGAGGGAGCG	62°C
Mcp9_Rv			GACGTCTGCCGTGCGGGGT	
Mcp10_Fw	Mecp2	Prom-ORI	ACCCCGCACGGCAGACGTC	68°C
Mcp10_Rv			GTGCAGCAGCACACAGGCTG	
Mcp11_Fw	Mecp2	Prom-ORI	GACCAGCCTGTGTGCTGCTG	66°C
Mcp11_Rv			CTCGACAAAGAGCAAGGGGTG	
Mcp12_Fw	Mecp2	Prom-ORI	CACCCCTTGCTCTTTGTGCGAG	66°C
Mcp12_Rv			CCAGCCTGGGCTCCACAAC	
Mcp13_Fw	Mecp2	Prom-ORI	GTTGTGGAGCCCAGGCTGG	66°C
Mcp13_Rv			CAATTGAGGGCGTCACCGCT	
Mcp14_Fw	Mecp2	Prom-ORI	AGCGGTGACGCCCTCAATTG	58°C
Mcp14_Rv			CCTCTTTTCCCTGCCTAAAC	
McpFl_3_Fw	Mecp2	Flanco 3'	GCATCCAATGCTCTTTGTGTC	60°C
McpFl_3_Rv			GTCTCTTGTTGAGCATTTGT	
AtpFl_5_Fw	Atp2b3	Flanco 5'	ACTTGATAACCTTTAGGGATCAG	58°C
AtpFl_5_Rv			CAGATACATTTCCCCAGAGG	
Atp1_Fw	Atp2b3	Prom-ORI	TCCCTCCGCGCTCAGAGCTG	66°C
Atp1_Rv			GACGCTGCGCGCTCGACTC	
Atp2_Fw	Atp2b3	Prom-ORI	GAGTCGAGCGCGCAGCGTC	66°C
Atp2_Rv			AGGGCTGTGGCCGAGTGCTC	
Atp3_Fw	Atp2b3	Prom-ORI	TGGGGAGCACTCGGCCACAG	66°C
Atp3_Rv			AGCAGTGCGGAGCAGGACCC	
Atp4_Fw	Atp2b3	Prom-ORI	GGGTCCTGCTCCGCACTGCT	64°C
Atp4_Rv			CCAAGGACGGAAGGCCGGG	
Atp5_Fw	Atp2b3	Prom-ORI	TCCGTCCTTGGAGCTGATCG	66°C
Atp5_Rv			GCTGTGGCTTTGCTGTTGC	
Atp6_Fw	Atp2b3	Prom-ORI	CGCAACAGCGAAAGCCACAGC	66°C
Atp6_Rv			GCACAACCTACCCCGGCATG	
AtpFl_M_Fw	Atp2b3	Flanco Medio	GCACGCGCATCTGTGTCTAG	64°C
AtpFl_M_Rv			CTATGCCTGCCTGGATCCCC	
AtpFl_3_Fw	Atp2b3	Flanco 3'	CACATGTTCTACTACTGTCC	58°C
AtpFl_3_Rv			ACTGTCCATGCTAGCGTCAG	

ANEXO I: TABLA DE OLIGONUCLEÓTIDOS UTILIZADOS

HstFl_5_Fw	Hs6st2	Flanco 5	GTAGTGTCCAGTCCTTGCGG	62°C
HstFl_5_Rv			TGAGTGAGTGTGGGGAAAGC	
Hst1_Fw	Hs6st2	Prom-ORI	TCTGGATGTGCAGGAACACG	64°C
Hst1_Rv			CGATCTCCTGCGCAAGGTAG	
Hst2_Fw	Hs6st2	Prom-ORI	CTACCTTGCGCAGGAGATCG	64°C
Hst2_Rv			TGAGAGTTCGGCCAGGTTTG	
Hst3_Fw	Hs6st2	Prom-ORI	ACAAACCTGGCCGAACCTCTC	66°C
Hst3_Rv			GTTTCAGCTCCCCGGTGCCG	
Hst4_Fw	Hs6st2	Prom-ORI	CGGCACCGGGGAGCTGAAC	66°C
Hst4_Rv			CACAGAATGCCAGCTCCTCC	
Hst5_Fw	Hs6st2	Prom-ORI	GGAGGAGCTGGCATTCTGTG	64°C
Hst5_Rv			GGTGATCGTCTCCAATACG	
Hst6_Fw	Hs6st2	Prom-ORI	CGTATTGGAGGACGATCACC	64°C
Hst6_Rv			ACAAGCTGCTGCTGGCTTTC	
Hst7_Fw	Hs6st2	Prom-ORI	CAAAGCCAGCAGCAGCTTGT	64°C
Hst7_Rv			CAGCGTCGGAACATGGATG	
Hst8_Fw	Hs6st2	Prom-ORI	ATCCATGTTCCCGACGCTGG	66°C
Hst8_Rv			GCGGGATTGACGTAACGAGC	
Hst9_Fw	Hs6st2	Prom-ORI	GCTCGTTACGTCAATCCCGC	64°C
Hst9_Rv			AGCGGGAGCGAGGCGAGC	
Hst10_Fw	Hs6st2	Prom-ORI	GCTCGCTCGCTCCCGCA	66°C
Hst10_Rv			GGAAAGGGGCGCAACGGTGG	
Hst11_Fw	Hs6st2	Prom-ORI	CCACCGTTGCGCCCCTTCC	66°C
Hst11_Rv			GCCGCCGTGGTGGAGTACT	
Hst12_Fw	Hs6st2	Prom-ORI	AGTACTCCACCACGGCGGC	64°C
Hst12_Rv			GAAGGAAGGAAGCGGGGAAG	
Hst13_Fw	Hs6st2	Prom-ORI	CTTCCCCGCTTCCTTCCTTC	64°C
Hst13_Rv			TGTGCGGCAGAAGCCACAGC	
Hst14_Fw	Hs6st2	Prom-ORI	GCTGTGGCTTCTGCCGACA	66°C
Hst14_Rv			AACGGCCAGCGGGATGGAG	
HstFl_3_Fw	Hs6st2	Flanco 3'	CTTTTCATCTGGGAAAACAG	58°C
HstFl_3_Rv			ATTAGCTAGGAAATGTCTGC	
PsFl_5_Fw	Pias3	Flanco 5'	ATTATTGATGGGTAGGGCTC	56°C
PsFl_5_Rv			GTAAGCCATAAAGAGGGGAG	
Ps1_Fw	Pias3	noProm-ORI	TATGGGCTGGATGGTGAGTG	60°C
Ps1_Rv			AAGGGCCAGCCATGACTTAG	
Ps2_Fw	Pias3	noProm-ORI	TGAGCTAAGTCATGGCTGGC	58°C
Ps2_Rv			GAGACCTGTGGGTGGTTAAG	
Ps3_Fw	Pias3	noProm-ORI	CTTAACCACCCACAGGTCTC	56°C
Ps3_Rv			CCCTCTTCTTACTCTCTGGC	
Ps4_Fw	Pias3	noProm-ORI	TCAGCCAGAGAGTAAGAAGAG	56°C
Ps4_Rv			CTGATGAGCTTTTCGATGGTC	
Ps5_Fw	Pias3	noProm-ORI	GACCATCGAAAGCTCATCAG	60°C
Ps5_Rv			TGACAGGGCAGTGCTTCTTG	
Ps6_Fw	Pias3	noProm-ORI	AAGAAGCACTGCCCTGTCAC	60°C
Ps6_Rv			TCATACCCTTTGCTTCCAGG	
Ps7_Fw	Pias3	noProm-ORI	AGCCCTTCTGGAAGCAAAG	58°C
Ps7_Rv			GGCAGTTCTGACTTCTGTCC	
Ps8_Fw	Pias3	noProm-ORI	GGGACAGAAGTCAGAACTGCC	62°C
Ps8_Rv			TCTTCCCTGTCCCAAAGCTC	
Ps9_Fw	Pias3	noProm-ORI	CGAGCTTTGGGACAGGGAAG	64°C
Ps9_Rv			GAGCTAGCCAAGCGCAGATG	
Ps10_Fw	Pias3	noProm-ORI	CTGCGCTTGGCTAGCTCTC	62°C
Ps10_Rv			AGGTCAGGGCTCTGTGAGTG	

ANEXO I: TABLA DE OLIGONUCLEÓTIDOS UTILIZADOS

Ps11_Fw	Pias3	noProm-ORI	CACTCACAGAGCCCTGACCTC	64°C
Ps11_Rv			TGCCAGTGTGCCATTG	
Ps12_Fw	Pias3	noProm-ORI	AATGGGCACACTGGGCAG	62°C
Ps12_Rv			AGGTGGTACTCATGTAGCGG	
Ps13_Fw	Pias3	noProm-ORI	GCTACATGAGTACCCACCTG	56°C
Ps13_Rv			GGTCAGTGTCTACCTTGG	
Ps14_Fw	Pias3	noProm-ORI	GCTGACATCCAAGGTAGAAC	56°C
Ps14_Rv			AAACAGGTCTGGAGAGGGTG	
PsFI_3_Fw	Pias3	Flanco 3'	TCATCTCTCCAGCACTACGG	60°C
PsFI_3_Rv			GGTGTCTCTCATCTAGCG	
SdFI_5_Fw	Syde1	Flanco 5'	CTTCCAGGAAGCCTTCTCAC	62°C
SdFI_5_Rv			AGAAGCAGTAAGGAGGGGAG	
Sd1_Fw	Syde1	noProm-ORI	TCTTAACTCCCCCTGGCTAC	62°C
Sd1_Rv			TGTCAAAGTAGCTGGCAGG	
Sd2_Fw	Syde1	noProm-ORI	CCCTGCCAGCTAGTTTGACA	58°C
Sd2_Rv			GGAGTGAGAAAGATGATTGG	
Sd3_Fw	Syde1	noProm-ORI	CCAATCATCTTTCTCACTCC	58°C
Sd3_Rv			TTAGCCCTCATGGAGCTCAG	
Sd4_Fw	Syde1	noProm-ORI	CTGAGCTCCATGAGGGCTA	62°C
Sd4_Rv			ACCCACGGATCATCTGTCC	
Sd5_Fw	Syde1	noProm-ORI	AAGGACAGATGATCCGTGGG	64°C
Sd5_Rv			CACTCCCTCCCCAGCATCA	
Sd6_Fw	Syde1	noProm-ORI	TGATGCTGGGGAGGGAGTG	60°C
Sd6_Rv			CTCCTTGAACCCCTTAAAC	
Sd7_Fw	Syde1	noProm-ORI	GTTTAAGGGGTTCCAAGGAG	62°C
Sd7_Rv			GTGAGCTGGCTGCCCCAC	
Sd8_Fw	Syde1	noProm-ORI	GTGGGGCAGCCAGCTCAC	64°C
Sd8_Rv			GAAAGTGTGTGGGGCAGATC	
Sd9_Fw	Syde1	noProm-ORI	GATCTGCCCCACACACTTTC	64°C
Sd9_Rv			CAGTCCCCAGGCCAGGTGC	
Sd10_Fw	Syde1	noProm-ORI	GCACCTGGCCTGGGGACTG	64°C
Sd10_Rv			CTGCAGCTTCTGGTAGAACG	
Sd11_Fw	Syde1	noProm-ORI	CGTTCTACCAGAAGCTGCAG	62°C
Sd11_Rv			AGCAGAGCCCCGTGTTTTTC	
Sd12_Fw	Syde1	noProm-ORI	GAAAACACGGGGCTCTGCT	62°C
Sd12_Rv			GCTAACATTGTCCGGAACAGC	
Sd13_Fw	Syde1	noProm-ORI	GCTGTTCCGACAATGTTAGC	62°C
Sd13_Rv			ACAAGTGGCCGTTCCGACTTG	
Sd14_Fw	Syde1	noProm-ORI	CAAGTCCAACGGCCAGTTGT	66°C
Sd14_Rv			GCCTCAGTCTCCTTGCATC	
Sd15_Fw	Syde1	noProm-ORI	AGATGCAAGGAGGACTGAGG	62°C
Sd15_Rv			GATGAGGGGAGTCAGGACAC	
SdFI_3_Fw	Syde1	Flanco 3'	GATAGCGGCTGATGACAGAG	58°C
SdFI_3_Rv			TAAGAAGTACTCGCACAGGC	
mNR_Fw	chrX	Región Negativa	CCAAACTTCAGCAGTTTTCC	58°C
mNR_Rv			GAAATTGTCTTGCAAGAGGA	

<i>Leishmania major</i>				
NOMBRE	CROM.	CARACTERÍSTICAS	SECUENCIA (5'→ 3')	T ^a
SSRc1_Fw	LmjF.28	SSRc (TTS-ORI)	TCTCCCCCTTAGGCCATGC	64
SSRc1_Rv			CCCTCCCCCTTTCACTTTA	
SSRc2_Fw	LmjF.21	SSRc (TTS-ORI)	GGTTCCTCGCCTTCCTTTC	64
SSRc2_Rv			GCAGTAGTGGTTGACAGCG	
HT1_Fw	LmjF.11	HT (TTS-ORI)	ATGGTCCGCACCAAGCAAG	64
HT1_Rv			GAAGAGGGTGGGATTTTTCGC	
HT2_Fw	LmjF.17	HT (TTS-ORI)	GGAATGCCCAAACACC	60
HT2_Rv			CAGCTACACAAGTGCGACG	
Int1_Fw	LmjF.20	IR (noTTS-ORI)	AGGAGATGCATCTCTCATTG	56
Int1_Rv			GTACAGGCAATTATACACACG	
Int2_Fw	LmjF.23	IR (noTTS-ORI)	AACAGGCGACAAGGAGAAAG	60
Int2_Rv			TCAGCACCTTCCTTTCTTCG	
Int3_Fw	LmjF.09	IR (noTTS-ORI)	GGAAACACGGAGGAGGGTAAC	60
Int3_Rv			TCTCTCTCACACATGCTCCAC	
FI1_Fw	LmjF.22	Flanco (no ORI)	GGAAGGGAGACGAAACCATC	64
FI1_Rv			GCGTTGACTTTCCGTTTCAG	
FI2_Fw	LmjF.04	Flanco (no ORI)	GACTTTGCCTTGACGAGTTG	60
FI2_Rv			CTACAAAGATGGTCCGAGCC	
FI3_Fw	LmjF.33	Flanco (no ORI)	GGGATTCACTACGCGCTG	60
FI3_Rv			CGTGAAGAAGAGTCGCTCC	
ImNR_Fw	LmjF.28	Región Negativa	CTCACCACCGATGTCTCAAC	60
ImNR_Rv			CACAAGTAACACAAAGGAAAAGGG	

ANEXO I: ZONAS EXCLUIDAS DEL ANÁLISIS

Crom.	Inicio	Fin	Crom.	Inicio	Fin	Crom.	Inicio	Fin
LmjF.01	0	1999	LmjF.14	246000	254000	LmjF.25	910845	912845
LmjF.01	266988	268988	LmjF.14	472800	474100	LmjF.26	0	1999
LmjF.02	0	1999	LmjF.14	480000	486500	LmjF.26	646500	675750
LmjF.02	270000	298000	LmjF.14	620644	622644	LmjF.26	878200	883750
LmjF.02	353712	355712	LmjF.15	0	1999	LmjF.26	1089540	1091540
LmjF.03	0	1999	LmjF.15	158500	196500	LmjF.27	0	1999
LmjF.03	66500	71000	LmjF.15	453000	471000	LmjF.27	56300	58700
LmjF.03	382502	384502	LmjF.15	598600	600800	LmjF.27	150000	154700
LmjF.04	0	1999	LmjF.15	627517	629517	LmjF.27	198900	203000
LmjF.04	47400	61600	LmjF.16	0	1999	LmjF.27	375000	383200
LmjF.04	103000	108000	LmjF.16	612000	616600	LmjF.27	782700	784800
LmjF.04	225450	230250	LmjF.16	702800	708800	LmjF.27	870300	871900
LmjF.04	470852	472852	LmjF.16	712651	714651	LmjF.27	987800	1061400
LmjF.05	0	1999	LmjF.17	0	6500	LmjF.27	1128424	1130424
LmjF.05	78700	80600	LmjF.17	28400	43400	LmjF.28	0	1999
LmjF.05	113500	118200	LmjF.17	682829	684829	LmjF.28	589400	591800
LmjF.05	447900	454250	LmjF.18	0	1999	LmjF.28	858200	859800
LmjF.05	463823	465823	LmjF.18	174000	185000	LmjF.28	1056900	1062900
LmjF.06	0	1999	LmjF.18	530000	551500	LmjF.28	1158104	1160104
LmjF.06	514869	516869	LmjF.18	737748	739748	LmjF.29	0	1999
LmjF.07	0	1999	LmjF.19	0	1999	LmjF.29	32800	35500
LmjF.07	594352	596352	LmjF.19	360400	370300	LmjF.29	424000	426600
LmjF.08	0	1999	LmjF.19	407750	408300	LmjF.29	791200	801500
LmjF.08	285000	390000	LmjF.19	700208	702208	LmjF.29	1210663	1212663
LmjF.08	452500	480000	LmjF.20	0	1999	LmjF.30	0	1999
LmjF.08	551500	556000	LmjF.20	322500	342200	LmjF.30	489300	509300
LmjF.08	572960	574960	LmjF.20	740537	742537	LmjF.30	525100	536100
LmjF.09	0	1999	LmjF.21	0	1999	LmjF.30	585200	609100
LmjF.09	64000	85000	LmjF.21	6000	9500	LmjF.30	1327300	1338400
LmjF.09	571434	573434	LmjF.21	575000	582500	LmjF.30	1399600	1403434
LmjF.10	0	1999	LmjF.21	770972	772972	LmjF.31	0	1999
LmjF.10	429200	435100	LmjF.22	0	1999	LmjF.31	148300	155800
LmjF.10	568865	570865	LmjF.22	546500	548500	LmjF.31	339800	369200
LmjF.11	0	1999	LmjF.22	557500	560100	LmjF.31	660800	671900
LmjF.11	17500	19200	LmjF.22	690500	716602	LmjF.31	1482328	1484328
LmjF.11	580573	582573	LmjF.23	0	1999	LmjF.32	0	1999
LmjF.12	0	1999	LmjF.23	37500	40500	LmjF.32	863900	864700
LmjF.12	376500	552500	LmjF.23	770565	772565	LmjF.32	892700	893800
LmjF.12	673346	675346	LmjF.24	0	1999	LmjF.32	1241000	1241900
LmjF.13	0	1999	LmjF.24	838950	840950	LmjF.32	1242800	1243400
LmjF.13	91500	119000	LmjF.25	0	1999	LmjF.32	1405500	1467200
LmjF.13	652595	654595	LmjF.25	617000	628000	LmjF.32	1602637	1604637
LmjF.14	0	1999	LmjF.25	902500	906500	LmjF.33	0	1999

ANEXO I: ZONAS EXCLUIDAS DEL ANÁLISIS

Crom.	Inicio	Fin	Crom.	Inicio	Fin	Crom.	Inicio	Fin
LmjF.33	89400	93300	LmjF.34	526700	531700	LmjF.35	1670800	1680900
LmjF.33	101900	103800	LmjF.34	706500	870000	LmjF.35	2088474	2090474
LmjF.33	112900	174600	LmjF.34	1141500	1150300	LmjF.36	0	1999
LmjF.33	307000	307900	LmjF.34	1172200	1180200	LmjF.36	5600	9400
LmjF.33	338200	395000	LmjF.34	1811900	1836500	LmjF.36	54600	61200
LmjF.33	676000	677350	LmjF.34	1864748	1866748	LmjF.36	466500	475100
LmjF.33	1024200	1025300	LmjF.35	0	1999	LmjF.36	863800	867300
LmjF.33	1304750	1334950	LmjF.35	4500	8100	LmjF.36	941400	943700
LmjF.33	1492800	1513000	LmjF.35	75600	77100	LmjF.36	984400	991100
LmjF.33	1581653	1583653	LmjF.35	159800	161500	LmjF.36	1370900	1373500
LmjF.34	0	1999	LmjF.35	167400	187400	LmjF.36	1880800	1886300
LmjF.34	304900	307000	LmjF.35	194400	245000	LmjF.36	2027180	2029200
LmjF.34	310900	317500	LmjF.35	256000	271300	LmjF.36	2416200	2425300

PUBLICACIONES QUE FORMAN PARTE DE ESTA TESIS DOCTORAL:

Lombraña R., Almeida R., Revuelta I., Madeira S., Herranz G., Bastolla U. y Gómez M. (2013) High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J* **32**: 2631-2644

Lombraña R., Almeida R., Álvarez A. y Gómez M. (2015) R-loops and initiation of DNA replication in human cells: a missing link? *Front Genet* doi:10.3389/fgene.2015.00158

Lombraña R., Álvarez A., Fernández-Justel JM., Almeida R., Poza-Carrión C., Gomes F., Calzada A., Requena JM. y Gómez M. (2016) Transcriptionally driven DNA replication programme of the human parasite *Leishmania major*. *Cell Rep* **16**: 1774-17786

OTRAS PUBLICACIONES:

Milligan L., Huynh-Thu VA., Delan-Forino C., Tuck A., Petfalski E., **Lombraña R.**, Sanguinetti G., Kudla G. y Tollervey D. (2016) Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* **12**:874