



VNIVERSITAT
DE VALÈNCIA (Q≈) Facultat de Ciències
Biològiques

Programa oficial de Postgrau en Biotecnología

BIOGRO: UN NUEVO MÉTODO DE ALTA RESOLUCIÓN PARA EL ESTUDIO DE LA TRANSCRIPCIÓN NACIENTE A ESCALA GENÓMICA EN LEVADURA

Trabajo realizado en el Departamento de Bioquímica y Biología Molecular
de la Universitat de València por Antonio Jordán Pla para optar al
grado de Doctor.

Valencia, Septiembre de 2013.

ÍNDICE GENERAL

ÍNDICE	I
ABREVIATURAS	IX
INTRODUCCIÓN	1
1.La levadura <i>Saccharomyces cerevisiae</i> como organismo modelo.	3
2.La transcripción en eucariotas.	4
2.1 Transcripción de la RNAPII.	4
2.1.1 La RNAPII y su dominio CTD.	5
2.1.2 Factores de transcripción y promotor mínimo.	8
2.1.3 Co-activadores y PIC.	10
2.1.4 Nucleosomas y remodeladores de cromatina.	12
2.1.5 Fases del ciclo transcripcional.	15
2.1.5.1 Iniciación.	15
2.1.5.2 Elongación.	16
2.1.5.3 Terminación y reiniciación.	18
2.1.6 Aspectos asociados a la transcripción.	23
2.1.6.1 Procesamiento de pre-mRNAs.	23
2.1.6.2 Coordinación entre síntesis y degradación de mRNAs.	25
3.Transcripción naciente.	29
3.1 Técnicas genómicas para la medición de tasas de transcripción.	30
3.2 Diferencias entre TR naciente y síntesis de mRNAs maduros.	32
3.3 Perfiles de densidad de RNAP como reflejo de diferencias en la tasa de elongación.	35
4.La transcripción de la RNAPIII.	36
4.1 Visión general de la estructura de la RNAPIII.	37
4.2 Regulación de la transcripción de la RNAPIII.	39

4.3 Organización espacial de la transcripción.	43
4.4 Biogénesis de los tRNAs.	44
4.5 Sesgo en el uso de codones y selección traduccional.	45
4.6 Elementos transponibles asociados a tRNAs.	46
OBJETIVOS	49
MATERIALES Y MÉTODOS	53
1.Materiales de especial relevancia.	55
2.Técnicas microbiológicas.	57
2.1 Cepas de microorganismos.	57
2.2 Medios y condiciones de cultivo.	57
3.Técnicas de biología molecular.	58
3.1 Obtención de ácidos nucleicos.	58
3.1.1 Preparación de RNasa A libre de DNasas.	58
3.1.2 Extracción de DNA de levadura.	58
3.1.3 Extracción de RNA de levadura.	59
3.1.3.1 Método del fenol ácido.	59
3.1.3.2 Método de la Proteinasa K.	60
3.1.4 Extracción de DNA plasmídico de <i>E. coli</i> .	61
3.2 Técnicas electroforéticas de separación de ácidos nucleicos.	61
3.2.1 Electroforesis nativa en gel de agarosa.	61
3.2.2 Electroforesis en gel desnaturizante de agarosa.	62
3.2.3 Electroforesis en gel desnaturizante de poliacrilamida.	62

3.2.4 Electroforesis capilar en dispositivos compactos.	62
3.3 Reacción en cadena de la polimerasa (PCR).	63
3.3.1 PCR convencional para generar moldes para IVT.	63
3.3.2 PCR cuantitativa.	64
3.3.3 Oligonucleótidos.	64
3.4 Transferencia de ácidos nucleicos a membranas.	65
3.4.1 <i>Dot Blot</i> .	65
3.4.2 <i>Northern Blot</i> a partir de geles de agarosa.	66
3.4.3 <i>Northern Blot</i> a partir de geles de poliacrilamida.	66
3.5 Biotinilación reversible de RNAs tiolados.	67
3.6 Aislamiento específico de RNAs marcados.	68
3.6.1 Inmunoprecipitación de RNA marcado con Bromo-U.	68
3.6.2 Inmunoprecipitación de RNA marcado con Metil-C.	70
3.6.3 Purificación de RNA biotinilado con bolas magnéticas.	71
3.7 Transcripción inversa (RT).	71
3.8 Transcripción <i>in vitro</i> (IVT).	72
3.9 Selección de ácidos nucleicos por tamaño.	72
3.10 Sonicación de DNA.	73
4. Técnicas genómicas.	74
4.1 <i>Genomic run-on</i> radiactivo (GRO).	74
4.2 <i>Genomic run-on</i> no radiactivo (BioGRO).	75
4.3 <i>Filter run-on</i> .	76

4.4 Hibridación de macrochips de DNA de nailon.	78
4.5 Hibridación de microchips de embaldosado.	78
4.6 Amplificación de la muestra de BioGRO.	79
5.Métodos de análisis bioinformático de los datos.	81
5.1 Análisis de imagen de los microchips de embaldosado.	81
5.2 Normalización de los datos.	82
5.2.1 Normalización de los datos mediante el software de Affymetrix.	82
5.2.2 Normalización de los datos mediante el lenguaje estadístico R.	82
5.3 Bases de datos y otras herramientas para el análisis bioinformático.	82
RESULTADOS	85
CAPÍTULO 1. Aspectos técnicos del método BioGRO.	87
1.1-Antecedentes de trabajo.	87
1.2-Depleción de mRNAs mediante RNasa A.	89
1.3-Aumento de la eficiencia del <i>run-on</i> tras congelación.	91
1.4-Incorporación de nucleótidos modificados mediante <i>run-on</i> tras RNasa.	94
1.4.1-Tio-UTP.	95
1.4.2-Bromo-UTP.	97
1.4.3-Metil-CTP.	98

1.5-GRO con biotina.	99
1.5.1. Análisis del tamaño de elongación con biotina en distintas cepas.	101
1.5.2. Selección de RNAs por tamaño.	101
1.6-Hibridación de la muestra.	104
1.7-Correlaciones de los datos de BioGRO con otros datos de TR de la literatura.	107
1.8- Corrección de datos de GRO y BioGRO por la tasa de <i>drop off</i> .	110
CAPÍTULO 2. Aplicación del método BioGRO a diferentes cepas de levadura.	
Estudio de cepas mutantes relacionadas con la síntesis y la degradación de RNAs.	115
2.1- <i>Gen Modelo</i> : perfil global de densidad de RNAPs activas; caso especial de los genes con intrones.	115
2.2-Análisis del metagen de los extremos 5' y 3' de los genes.	118
-2.2.1. Perfil de CHIP de RNAPII en los extremos 5' y 3' de los genes en relación al perfil de RNAPs elongantes.	121
-2.2.2. Análisis del perfil promedio de RNAPs activas en ciertos subgrupos de genes.	124
-2.2.3. Análisis de la presencia de RNAPs activas más allá de los límites canónicos de los transcritos.	128
2.3-Estudio del nascentoma del mutante <i>rpb4Δ</i> .	132
2.4-Estudio del nascentoma del mutante <i>xrn1Δ</i> .	134
CAPITULO 3. Relación entre RNAPs activas y cromatina.	137
3.1-Análisis de datos de nucleosomas.	137

3.2-Análisis de la cepa mutante <i>isw2Δ</i> .	140
3.3- Análisis de la cepa mutante <i>dst1Δ</i> .	147
CAPÍTULO 4. Caracterización del nascentoma de la RNAPIII.	152
4.1-Visión general de la transcripción nascente de la RNAPIII. Estudio de los genes de tRNAs.	152
4.2-Estudio de los genes de tRNAs con intrones.	155
4.3-Selección traduccional y transcripción nascente.	157
4.4-El nascentoma de RNAPIII en las cepas mutantes <i>dst12Δ</i> y <i>isw2Δ</i> .	159
4.5-Los genes de tRNAs como puntos calientes de inserción. de elementos transponibles. Relación con la nTR.	162
DISCUSIÓN GENERAL	165
1. Validación de la autenticidad del RNA nascente.	167
2. Ventajas del GRO frente al GRO <i>clásico</i> y a otras técnicas de RNA nascente.	172
3. Modelo de la autopista.	174
4. Características generales del gen <i>tipo</i> de RNAPII. Subtipos.	176
5. Estudio del nascentoma de los mutantes <i>rpb4Δ</i> y <i>xrn1Δ</i> .	181
6. RNAPII y nucleosomas, interdependencia.	183
7. Nascentoma de RNAPIII.	185
8. Posibilidades de mejora de la técnica y aplicación a otras levaduras.	188
CONCLUSIONES	191
BIBLIOGRAFÍA	197

ABREVIATURAS

BrEt - Bromuro de etidio

cDNA - DNA copia producto de una retrotranscripción

ChIP - Inmunoprecipitación de cromatina

cRNA - RNA copia producto de una transcripción *in vitro*

CTD - Dominio carboxi-terminal de la RNA Polimerasa

DO - Densidad óptica

dpm - Desintegraciones por minuto

gDNA - DNA genómico

GRO - *Genomic Run-on*

GTF - Factores generales de transcripción

IP - Inmunoprecipitación

IVT - Transcripción *In vitro*

kb - Kilobase

min - Minutos

mRNA - RNA mensajero

NFR - Región libre de nucleosomas

nRNA - RNA naciente

nt - Nucleótido

nTR - Tasa de transcripción naciente

ORF - Pauta abierta de lectura

p-valor - Nivel de significación α más pequeño que permite rechazar la hipótesis nula

pA - Sitio de corte y poliadenilación

pb - Pares de bases

PCR - Reacción en cadena de la polimerasa

PIC - Complejo de preiniciación

qPCR - PCR cuantitativa

RA - Cantidad de mRNA

RiBi - Genes de Biogénesis de Ribosomas

RNA - Ácido ribonucleico

RNAP - RNA polimerasa

RP - Genes de proteínas ribosómicas

RPCC - RNA Pol CHIP-on-chip

rRNA - RNA ribosómico

rpm - Revoluciones por minuto

RT - Retrotranscripción

Sarkosil - Lauroilsarcosinato de sodio

Seg - Segundos

Ser - Serina

SR - Tasa de síntesis de mRNA maduro

TR - Tasa de transcripción

tRNA - RNA de transferencia

TSS - Sitio de inicio de la transcripción

TTS - Sitio de terminación de la transcripción

UTP- Uridina trifosfato

UTR - Región no traducida de los mRNAs

WT - Cepa silvestre

INTRODUCCIÓN

1. La levadura *Saccharomyces cerevisiae* como organismo modelo

Las levaduras son hongos levaduriformes unicelulares que, a diferencia de los hongos filamentosos, carecen de distinción entre cuerpos vegetativo y reproductivo. Se clasifican, junto con mohos y setas, dentro del reino *Fungi*, que está formando por dos *phyla* diferentes: Ascomycetes y Basidiomycetes. El ciclo de vida de las levaduras alterna dos formas, una haploide y otra diploide. Ambas formas se reproducen de forma asexual por gemación. En condiciones determinadas la forma diploide es capaz de reproducirse sexualmente. En estos casos se produce la meiosis en la célula formándose un asca que contiene ascosporas haploides. A pesar de que las levaduras son organismos unicelulares, poseen una organización celular similar a la de eucariotas superiores, incluidos los humanos. Una de las especies más representativas de levadura, perteneciente al *phylum* Ascomycetes, es *Saccharomyces cerevisiae*. Su tamaño es de unas 4 micras de diámetro, y contiene un genoma haploide de 12 megabases, formado por 16 cromosomas y ~6000 genes (Goffeau *et al.*, 1996). El presente trabajo de tesis doctoral es el resultado de estudios globales de la transcripción, tomando como modelo a la mencionada especie de levadura.

S. cerevisiae ha formado parte de la vida de los humanos desde hace miles de años. Su valor ha radicado históricamente en sus importantes utilidades industriales, como la producción de cerveza, pan y vino (Pretorius, 2000). De manera más reciente se ha convertido en una poderosa herramienta desde el punto de vista científico, empleándose como modelo simple de célula eucariota, ya que la maquinaria molecular de muchos procesos celulares se encuentra conservada tanto en levaduras como en plantas y en mamíferos. Esto, junto a otras ventajas técnicas, tales como su facilidad de cultivo, su rápido crecimiento (aproximadamente dos horas para completar un ciclo celular) y el conocimiento de la secuencia completa de su genoma (Goffeau *et al.*, 1996), la han convertido en una importante herramienta en estudios de Biología Molecular y tecnologías ómicas. *S. cerevisiae* es un organismo modelo idóneo para el desarrollo de nuevas técnicas, y proporciona un punto de partida para el análisis de organismos eucariotas más complejos.

2. La transcripción en eucariotas

La transcripción del DNA es el primer proceso de la expresión génica, mediante el cual se transfiere la información contenida en la secuencia del DNA hacia la secuencia de proteína utilizando diversos RNAs como intermediarios informativos y funcionales. Durante la transcripción, las secuencias codificantes del DNA son copiadas a RNA mediante enzimas llamadas RNA Polimerasas (RNAPs) que sintetizan RNAs mensajeros que mantienen la información de la secuencia del DNA. De esta manera, la transcripción de la parte codificante del DNA también podría llamarse síntesis del RNA mensajero (mRNAs).

La transcripción es llevada a cabo por RNAPs, enzimas capaces de catalizar la síntesis de una molécula de RNA a partir de nucleótidos trifosfato (NTPs). En el caso de los eucariotas, el proceso se realiza en el núcleo, y es similar al de las procariotas, pero de mayor complejidad. Existen tres RNAPs en eucariotas: la RNAPI lleva a cabo la transcripción de los RNAs ribosómicos (rRNAs) 25S, 18S y 5,8S. La RNAPII se encarga de la transcripción de los mRNAs que posteriormente serán traducidos para dar lugar a proteínas, así como de algunos RNAs nucleares pequeños (snRNA) y otros RNAs no codificantes. Finalmente, la RNAPIII transcribe los RNA de transferencia (tRNAs), el RNA ribosómico 5S y algunos snRNAs. Las tres polimerasas poseen una estructura similar, presentando algunas subunidades comunes entre ellas, y otras específicas (Sentenac, 1985).

2.1 Transcripción de la RNAPII

La expresión génica es un proceso compuesto de muchos niveles, y cada uno de ellos está sujeto a una estricta regulación. La maquinaria transcripcional la componen tanto la propia enzima sintetizadora del RNA naciente, en este caso la RNAPII, como toda una plétora de elementos y factores asociados: elementos de secuencia del DNA de las zonas promotoras, factores de transcripción, co-activadores, nucleosomas y sus remodeladores, y factores que participan en la maduración del transcrito naciente.

Los siguientes apartados se encargan de revisar el conocimiento que se tiene de cada uno de estos elementos y cómo afectan al proceso transcripcional en levadura.

2.1.1 La RNAPII y su dominio CTD

La RNAPII es un complejo multiproteico muy conservado en eucariotas compuesto de doce subunidades (Cramer, 2004) codificadas por los genes Rpb1 a 12. A excepción de Rpb4 y Rpb9, todas estas subunidades son esenciales para la célula (Woychik y Young, 1989; Woychik *et al.*, 1991) y son denominadas el holoenzima. Rpb4 y Rpb7 forman un subcomplejo heterodimérico dissociable del núcleo catalítico (Edwards *et al.*, 1991; Gnatt *et al.*, 2001; Cramer *et al.*, 2001; Armache *et al.*, 2005). Rpb1 y Rpb2 son las subunidades más conservadas y de mayor tamaño. Su asociación forma el sitio activo. El sitio activo contiene en su interior un ión Mg^{2+} como cofactor, alrededor del cual se asocian el resto de subunidades para formar el holoenzima. El sitio activo es el lugar donde se produce la biogénesis del RNA naciente a partir de un DNA que actúa como molde. Existen multitud de factores capaces de interactuar con la RNAPII y modular la transcripción.

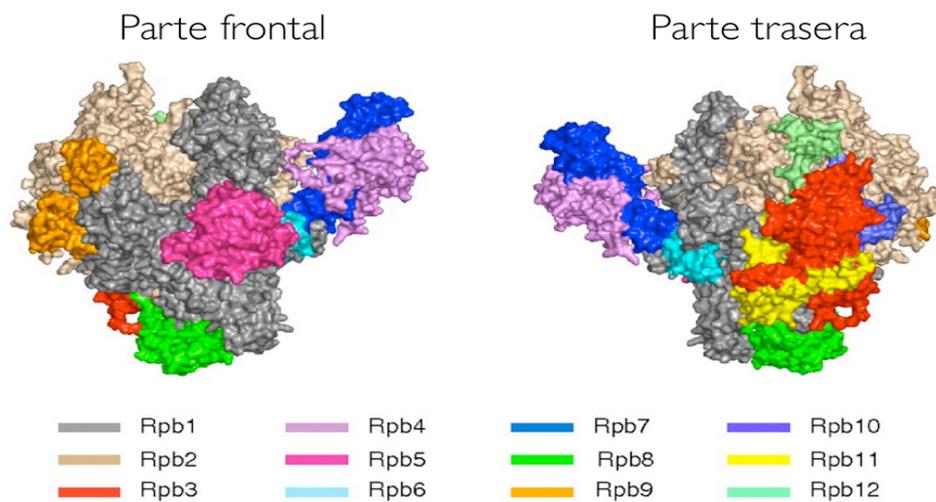


Figura i1. Estructura del holoenzima de la RNAPII. Adaptado de Spåhr *et al.* (2009).

La unidad más grande de la RNAPII, Rpb1, exhibe un dominio C-terminal (dominio CTD, *Carboxy-Terminal Domain*), consistente en repeticiones en tándem del heptapéptido YSPTSPS (26 en de *S. cerevisiae*, 52 en humanos). Este dominio, esencial para la viabilidad, actúa a modo de andamio dinámico para la unión de proteínas que interactúan con el complejo de transcripción y llevan a cabo funciones asociadas a la transcripción. Entre tales funciones se incluyen la modificación de histonas (Phatnani y Greenleaf, 2006), el procesamiento del pre-mRNA (formación de la caperuza 5', de la cola poliA en 3', y eliminación de intrones), y su exportación al citoplasma (Hahn y Young, 2011).

El dominio CTD está situado estratégicamente cerca del canal de salida del RNA naciente, y recibe toda una serie de modificaciones post-traduccionales a lo largo del ciclo transcripcional, entre las que se encuentran la glicosilación, la isomerización de las prolinas y la fosforilación de las serinas en posiciones 2, 5 y 7 (Ser2, Ser5, Ser7) (Meinhart *et al.*, 2005). El caso concreto de las modificaciones por fosforilación y desfosforilación de las tres serinas distintas de cada repetición del heptapéptido exhibe un gran potencial combinatorio, lo que ha llevado a postular la existencia de un “código CTD” que es necesario para organizar el reclutamiento secuencial de los numerosos factores involucrados en las distintas fases de iniciación, elongación y terminación de la transcripción (Buratowski, 2003; Egloff *et al.*, 2012; Bataille *et al.*, 2012).

La fosforilación del CTD la realizan principalmente los miembros de la familia de ciclinas dependientes de quinasa (CDK, *Cyclin-Dependent Kinase*), compuestas habitualmente por una subunidad catalítica y una subunidad ciclina reguladora (Meinhart *et al.*, 2005). En *S. cerevisiae*, al menos cuatro complejos Cdk son capaces de fosforilar el CTD (Kin28-Ccl1-Tfb3, Srb10-Srb11, Ctk1-Ctk2-Ctk3, y Bur1-Bur2), y todos ellos tienen un papel en la regulación de la transcripción (Phatnani y Greenleaf, 2006).

Existen diversos factores capaces de fosforilar y modular cada uno de los tipos de fosforilación de la RNAPII durante las distintas fases del proceso de transcripción (Buratowski, 2009; Bataille *et al.*, 2012). En cuanto a la desfosforilación del CTD, se conocen tres fosfatasa implicadas: Fcp1 tiene un papel positivo en la fase de elongación (Cho *et al.*, 1999; Mandal *et al.*, 2002), y es capaz de desfosforilar tanto la Ser2 como la Ser5 en humanos, aunque en levaduras presenta una preferencia por la Ser2 (Cho *et al.*, 2001; Hausmann y Shuman, 2002; Lin *et al.*, 2002). Ssu72, desfosforila los residuos de

Ser5 y juega un papel en la terminación y el corte del extremo 3' del pre-mRNA (Krishnamurthy *et al.*, 2004). Se ha descrito una tercera fosfatasa recientemente, Rtr1, que juega un papel en la transición del estado de fosforilación en Ser5 al de Ser2 (Mosley *et al.*, 2009). Al inicio de la transcripción, la RNAPII forma el complejo de pre-iniciación (PIC, *Pre-Initiation Complex*) al ser reclutada en forma hipofosforilada a la zona del promotor por los factores generales de transcripción (Lu *et al.*, 1991). Después de unirse al promotor, la Ser5 es rápidamente fosforilada por Kin28 (Cdk7 en humanos), la subunidad quinasa del factor general de transcripción TFIIH. Este evento facilita la disociación de la RNAPII del PIC, favoreciendo que ésta escape del promotor (Saunders *et al.*, 2006, Max *et al.*, 2007). De manera concomitante, la fosforilación de la Ser5 promueve el reclutamiento de los factores de formación de la caperuza y la escisión de intrones (Perales y Bentley, 2009), además del complejo metiltransferasa Set1 (COMPASS en levadura, MLL en humanos) y los complejos histona-desacetilasa Set3C y Rpd3 (Drouin *et al.*, 2010; Kim y Buratowski, 2009; Ng *et al.*, 2003). Durante la elongación temprana, los niveles de fosforilación de Ser5 disminuyen, mientras que la Ser2 comienza a ser fosforilada por las quinasas Ctk1 y/o Bur1 (Bartkowiak *et al.*, 2010; Keogh *et al.*, 2003; Komarnitski *et al.*, 2000), lo que lleva al reclutamiento de la histona-metiltransferasa Set2 y a la activación del complejo Rpd3, que bloquea la transcripción críptica en el interior de los genes (Carrozza *et al.*, 2005). Cuando la RNAPII alcanza el sitio de terminación, factores de procesamiento del mRNA pertenecientes al complejo CPF (*Cleavage and Polyadenylation Factor*) se unen al CTD para llevar a cabo las reacciones de corte y poliadenilación del RNA naciente (Lunde *et al.*, 2010). Este proceso también requiere la apropiada fosforilación del CTD. Durante esta fase el CTD es desfosforilado, y la RNAPII se recicla para una nueva ronda de transcripción (Ansari y Hampsey, 2005). Por su parte, la Ser7 parece que se fosforila en las regiones 5' de los genes y esta asociada a la transcripción de algunos grupos funcionales específicos de RNAs como los snRNAs (Egloff *et al.*, 2007), aunque su función en los genes codificantes de proteínas todavía no está del todo clara.

2.1.2 Factores de transcripción y promotor mínimo.

Los factores de transcripción (TF, *Transcription Factor*) conectan los procesos celulares a la transcripción. Activados por varios tipos de eventos, como fosforilación, oxidación, cambio de localización del citoplasma al núcleo, proteólisis o interacción con proteínas reguladoras, los TFs son capaces de iniciar (activadores) o reprimir (represores) la transcripción.

Las actividades de los TFs se organizan en módulos funcionales: el módulo de unión al DNA, que reconoce secuencias específicas del DNA, el módulo de transactivación, con potencial para iniciar la transcripción, y el módulo de dimerización, que media interacciones homo- y diméricas (Kadonaga, 2004). Estos módulos funcionales pueden aparecer combinados en una sola proteína, o pueden estar compartidos entre varias que actúan de manera sinérgica (Brivanlou y Darnell, 2002).

Un mecanismo principal de activación de la transcripción es el reclutamiento de proteínas reguladoras a la región promotora de genes específicos. Muchas de estas proteínas son factores asociados a la cromatina, como remodeladores, histona-acetiltransferasas, metilasas, histona desacetilasas o quinasas, que promueven la actividad transcripcional.

Los TFs represores se encargan de prevenir que la maquinaria de transcripción empiece su actividad. Pueden ser globales o específicos de genes, y actuar a corta o larga distancia del sitio de inicio de la transcripción, TSS (*Transcription Start Site*) (Mannervick *et al.*, 1999). Para ejercer la represión, los TFs pueden utilizar varios mecanismos: 1) disminuir la concentración de activadores/co-activadores en la zona del promotor, 2) contrarrestar el efecto estimulador de la transcripción de estas proteínas, 3) competir con componentes del PIC por el contacto con el DNA de la zona del promotor, o 4) competir con los remodeladores de cromatina (Gaston y Jayaraman, 2003).

La expresión de la mayoría de los genes eucarióticos depende del contexto, y puede ser modulada mediante la actividad combinada de grupos distintos de TFs. Así, las regiones promotoras contienen muchos sitios de unión de múltiples TFs, lo cual hace posible que cada gen pueda ser regulado por múltiples rutas de señalización (Orphanides y Reinberg,

2002). En algunos casos, un mismo TF puede tener funciones de activador o de represor. Es el caso por ejemplo del heterodímero Bur6-Ydr1 (NC2), un factor que interactúa con el promotor y componentes de la maquinaria basal de transcripción, como la proteína de unión a la caja TATA (TBP), y que se considera tanto un represor como un activador de la transcripción (Cang y Prelich, 2002; Peiró-Chová y Estruch, 2007).

Los TFs se denominan *generales*, GTFs (*General Transcription Factor*) cuando afectan a la expresión de la mayoría de genes. Los GTFs y la RNAPII se ensamblan al PIC en elementos con secuencias específicas del promotor. El conjunto mínimo de elementos requeridos para producir un nivel de transcripción basal se conoce como el promotor mínimo o basal (Smale y Kadonaga, 2003). La arquitectura del promotor mínimo eucariota consiste en la caja TATA, el iniciador (INR, *INITIATOR*), el elemento aguas abajo (DPE, *Downstream Promoter Element*), el elemento motivo-10 (MTE, *Motif-Ten Element*), y el elemento de reconocimiento de TFIIB (BRE, *TFIIB Recognition Element*). Estos elementos son reconocidos por los GTFs y sirven como plataforma para el ensamblaje del PIC. Cabe destacar que en *S. cerevisiae* la arquitectura descrita es la más general, pero no la única, ya que el elemento DPE es sustituido por otro elemento, y que hay algo más del 10% de los genes que posee una secuencia TATA canónica (el resto posee una secuencia similar, conocida como secuencia TATA-like. Rhee y Pugh, 2012).

Elemento TATA: El elemento TATA, o caja TATA, consiste en una secuencia rica en A/T con un dominio conservado TATAWAWR (W indica A/T, R indica A/G), que es reconocido por la proteína de unión a TATA, TBP (*TATA Binding Protein*) (Basehoar *et al.*, 2004). En *S. cerevisiae*, la caja TATA se localiza 50-120 pb aguas arriba del TSS (Zhang y Dietrich, 2005). La TBP reconoce la secuencia TATA en sus extremos 5' y 3', y desenrolla parcialmente la hélice del DNA (Smale y Kadonaga, 2003), lo que ayuda a nuclear el PIC. La TBP ha sido descrita como una subunidad del GTF TFIID, e interacciona también con el complejo acetiltransferasa Spt-ADA-Gcn5 (complejo SAGA). Dependiendo de la secuencia que la TBP encuentre en el promotor, interaccionará con TFIID o con SAGA. Los promotores con una secuencia TATA consenso son el 13% del total, dependen de SAGA para su activación y generalmente son de respuesta a estrés (Hahn y Young, 2011). El promotor del resto de genes, conocidos como *TATA-less*, ha sido considerado hasta hace poco carente de secuencia o elemento TATA. No obstante, estudios recientes han sido capaces de

identificar elementos de secuencia a los que se une TBP en los promotores TATA-Less: los elementos GAE (Seizl *et al.*, 2011) y los elementos *TATA-like* (Rhee y Pugh, 2012).

Iniciador (INR): La secuencia iniciadora se localiza alrededor del TSS, y ayuda a la unión de TBP, por lo que es funcionalmente parecida al elemento TATA. Puede ser unida de manera débil por la RNAPII, o fuertemente cuando se une a ella el conjunto de RNAPII, TFIIB, TFIID y TFIIF (Dikstein, 2011).

BRE: El elemento reconocido por TFIIB (*TFIIB Recognition Element*) es una secuencia que se encuentra en los extremos de la caja TATA a la que se une TFIIB

2.1.3 Co-activadores y PIC

En levadura existen dos mecanismos principales de activación transcripcional, mediados o bien por SAGA, o por TFIID. Ambos complejos comparten a la TBP y están involucrados en la formación del PIC.

SAGA: El complejo remodelador de histonas SAGA de *S. cerevisiae* está formado por 21 subunidades, 15 esenciales y 6 no esenciales, que regulan la transcripción mediante la modificación covalente de las histonas (Bhaumik, 2011).

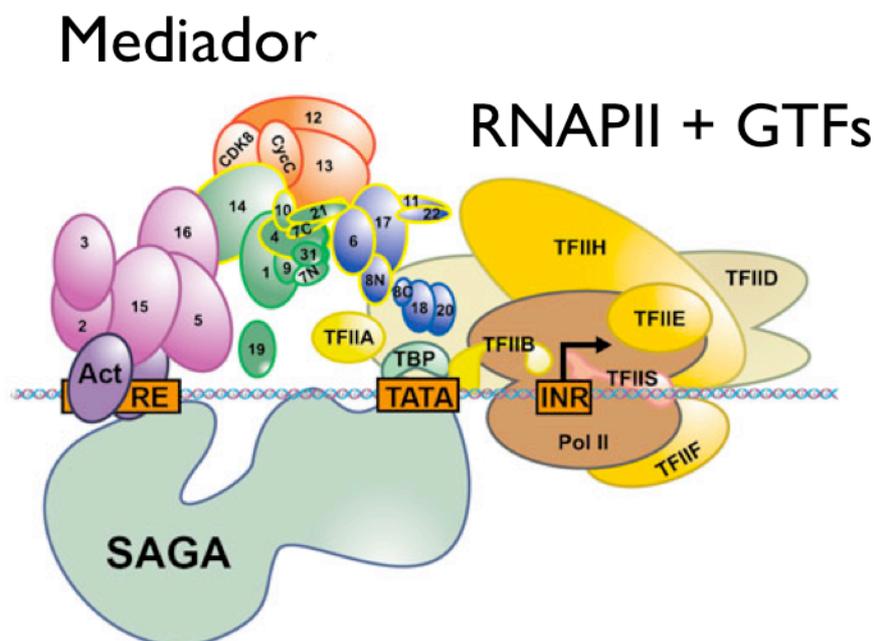


Figura i2. Modelo estructural del complejo de promotor mínimo y co-activadores. Tomado de Seizl (2011).

Los componentes Gcn5 y Ubp8 poseen actividades histona acetiltransferasa (HAT, *Histone Acetil Transferase*), y histona deubiquitinasas, respectivamente. Además, SAGA es diana de activadores transcripcionales que promueven su reclutamiento al promotor, sus actividades de modificación de histonas y su unión estable a TBP en el promotor (Daniel y Grant, 2007; Bhaumik, 2011).

TFIID: TFIID está compuesto de TBP y un conjunto conservado de 14 factores asociados a TBP, TAFs (*TBP Associated Factors*). Como ya se ha comentado, el 87% de los genes requiere TFIID para su activación, y este mecanismo es independiente de SAGA.

Como hemos visto hasta ahora, para que se produzca la iniciación de la transcripción, los TFs, co-activadores y RNAPII tienen que asociarse en la región promotora. Esta asociación se conoce como PIC, y sus componentes, con la ayuda de muchos otros, regulan la actividad de la RNAPII a lo largo de los pasos centrales del ciclo transcripcional.

Mediador: el complejo Mediador hace de puente entre los TFs y la maquinaria basal de la RNAPII, estimulando así la transcripción dependiente de TFs activadores. El complejo en *S. cerevisiae* está formado por 25 subunidades que pueden subdividirse en cuatro módulos funcionales distintos:

- El módulo de cabeza (*head module*) consiste de siete subunidades individuales (Med6, Med8, Med11, Med17, Med18, Med20 y Med22). Este módulo se une de manera débil al complejo de TBP, RNAPII, TFIIB, TFIIF y el DNA del promotor, con lo que se le atribuye una función estabilizadora de intermedios de ensamblaje del PIC.

- El módulo medio (*middle module*) es un subcomplejo de nueve subunidades (Med1, Med4, Med7, Med9, Med10, Med14, Med19, Med21 y Med31) diana de TFs específicos de gen.

- El módulo de cola (*tail module*) interacciona con una gran variedad de TFs activadores y consiste de las subunidades Med2, Med3, Med5, Med15 y Med16.

- El módulo quinasa (*kinase module*) consiste en cuatro subunidades disociables del Mediador (Med12, Med13, Srb10 y Srb11), que tienen efectos positivos y negativos en la expresión génica (Taatjes, 2010). El par Srb10/Srb11 (CDK8/ ciclina

C) fosforila la Ser5 del CTD prematuramente, evitando así la formación del PIC y reprimiendo la transcripción. Por otra parte, la subunidad Kin28 de TFIIF es la encargada de fosforilar el CTD, disociando parcialmente el Mediador del PIC y permitiendo la transcripción (Svejstrup *et al.*, 1997). Esta actividad de TFIIF está bajo el control del Mediador (Max *et al.*, 2007; Sikorski y Buratowski, 2009).

2.1.4 Nucleosomas y remodeladores de cromatina.

La transcripción de la RNAPII ocurre en un contexto cromatínico en las células eucarióticas. La cromatina es generalmente inhibitoria de la transcripción, por lo que se necesita toda una variedad de mecanismos para activar la transcripción de un molde de DNA cubierto de nucleosomas. Uno de los primeros pasos es la interacción de grandes complejos co-activadores con proteínas activadoras pequeñas con el fin de identificar los promotores de los genes preparados para ser transcritos. Los complejos remodeladores de cromatina que utilizan energía derivada de la hidrólisis del ATP para mover o desplazar a los nucleosomas del DNA facilitan el reclutamiento y el ensamblaje de estos complejos en el promotor, permitiendo una activación génica rápida. Durante la elongación de la transcripción los nucleosomas deben ser evacuados para que el paso de la RNAPII sea eficiente. Además, esos mismos nucleosomas deben ser rápidamente reensamblados y modificados posteriormente para evitar iniciaciones inapropiadas desde elementos funcionalmente parecidos a promotores dentro de las regiones codificantes (Weake y Workman, 2011).

Como se ha dicho, los genomas eucarióticos se empaquetan en la cromatina, cuya unidad básica de repetición es el nucleosoma. El nucleosoma consiste en un octámero de histonas envuelto por 147 pb del DNA (Richmond y Davey, 2003). Los nucleosomas se disponen en series espaciadas regularmente, a modo de cuentas de collar, siendo la longitud de la región espaciadora variable entre especies y tipos celulares. En levadura, el espaciado promedio entre nucleosomas es de 165 pb, y la longitud de la región espaciadora es de aproximadamente 18 pb. A pesar de que inicialmente se creía que los nucleosomas cubrían todo el DNA genómico de manera inespecífica, hace mucho tiempo

que se sabe que ocupan posiciones preferentes a lo largo del genoma (Pérez-Ortín *et al.*, 1989). Estudios a escala genómica de alta resolución han revelado un patrón común: los nucleosomas están ausentes de muchas (aunque no de todas) las regiones intensificadoras, promotoras y terminadoras, y ocupan posiciones distintas en regiones génicas e intergénicas (Yuan *et al.*, 2005; Lee *et al.*, 2007; Mavrich *et al.*, 2008). Así, se conoce la existencia de las regiones de 150-200 pb libres de nucleosomas (NFRs, *Nucleosome Free Regions*) inmediatamente arriba del TSS, e inmediatamente aguas abajo del sitio de corte y poliadenilación en la región 3' (en adelante, sitio pA). En levadura, los nucleosomas que flanquean el TSS a izquierda y derecha se denominan -1 y +1, y se localizan en posiciones muy concretas, mientras que el posicionamiento de los nucleosomas subsiguientes se hace gradualmente más difuso hacia la región 3' del transcrito (Yassour *et al.*, 2008; Brogaard *et al.*, 2012).

La disposición de los nucleosomas sobre el DNA afecta un amplio abanico de procesos biológicos, incluyendo la actividad de la RNAP, la unión de los TFs, la replicación del DNA, la estructura de los centrómeros y el procesado de intrones (Rando y Winston, 2012). En relación a los TFs, se sabe que muchos de sus efectos son mediados a través de interacciones con reguladores y co-reguladores que cambian el estado de la cromatina a para hacerla más abierta (activación de la transcripción), o más cerrada (represión) (Koerber *et al.*, 2009). De la misma manera, la disrupción del patrón general de nucleosomas puede tener toda una serie de consecuencias a nivel celular, una de las cuales es la aparición de un aumento en la transcripción críptica, como resultado de un desenmascarado de secuencias parecidas a promotores, o de la alteración de modificaciones covalentes de las histonas que afectan a la función de factores accesorios encargados de la degradación de estos transcritos (Terzi *et al.*, 2011). Estas modificaciones se producen en las colas N-terminal de las histonas, y tienen dos consecuencias principales: alterar la carga neta de los nucleosomas de manera que los contactos entre estos y el DNA se hagan más débiles (acetilaciones, que promueven el acceso al DNA de otras proteínas no histonas), o más fuertes (metilaciones, que cierran el paso a otras proteínas).

Los nucleosomas distorsionan y ocluyen el acceso al DNA genómico de la mayoría de las proteínas de unión al DNA, y sus posiciones exactas afectan a la estructura de la fibra de

cromatina (Richmond y Davey, 2003). Cambios de incluso un par de bases pueden afectar a la estructura de la cromatina y la cinética de las uniones de proteínas (Raveh-Sadka *et al.*, 2009; Mao *et al.*, 2011). Mucho se ha hablado sobre los determinantes del posicionamiento de los nucleosomas del DNA, y la conclusión a la que se ha llegado hasta el momento es que existen tres factores fundamentales: 1) la propia secuencia del DNA (Segal y Widom, 2009), 2) la actuación de factores remodeladores que mueven enzimáticamente las posiciones de los nucleosomas (Yen *et al.*, 2012), y 3) la propia actividad transcripcional de la RNAPII (Jones, 2012).

Los remodeladores propagan el movimiento de los nucleosomas mediante alteraciones de los contactos histona-DNA dependientes del consumo de ATP (Peterson y Workman, 2000; Zhang *et al.*, 2011). Estos movimientos comprenden el deslizamiento, la eyección, la inserción o la reestructuración de los nucleosomas o de las histonas. (Mohrman y Verrijzer, 2005; Vignali *et al.*, 2000; van Vugt *et al.*, 2007). Los remodeladores se clasifican en base a la homología de secuencia de la subunidad ATPasa, pudiéndose subdividir en cuatro grandes familias: ISWI (ISW1a, ISW1b, y ISW2), INO80/SWR1, CHD, y SWI/SNF (incluyendo RSC) (Mellor y Morillon, 2004; Hota y Bartholomew, 2011). No obstante, se pueden resumir en dos grupos atendiendo al mecanismo bioquímico que utilizan para movilizar los nucleosomas: el grupo ISWI y el grupo SWI/SNF, según necesiten al DNA espaciador o no. Dentro del grupo ISWI encontramos a Isw1, Isw2, Chd1, e Ino80, que dependen del DNA espaciador (Whitehouse *et al.*, 2007; Hota y Bartholomew, 2011). *In vitro*, el grupo ISWI mueve los nucleosomas desde el final de un fragmento de DNA hasta el medio, actividad que podría ayudar *in vivo* a espaciar los nucleosomas regularmente en una serie (Tsukiyama *et al.*, 1999; Udagama *et al.*, 2011). Por su parte, SWI/SNF y RSC transfieren octámeros de histonas a un DNA desnudo *in trans*, y además mueven lateralmente los nucleosomas, pero sin necesidad de DNA espaciador (Whitehouse *et al.*, 1999; Kassabov *et al.*, 2003). Otra función de RSC es la de eliminar nucleosomas del DNA, mediante un mecanismo llamado desalojo de nucleosomas (*nucleosome eviction*) (Montel *et al.*, 2011).

El papel a escala genómica de los remodeladores en levadura ha sido investigado mediante diversas mutaciones, descubriéndose por ejemplo que Isw2 se une a los nucleosomas que flanquean los NFRs y los acerca a ellos para reprimir la transcripción

críptica (Whitehouse *et al.*, 2007; Yen y Pugh, 2012); que, por el contrario RSC los mueve alejándolos del NFR (Hartley y Madhani, 2009); y que Isw1 funciona, de manera colaborativa con Chd1, más hacia la porción media del transcrito, donde promueve el reposicionamiento orientando los nucleosomas hacia el extremo 3' (Tirosh *et al.*, 2010). La delección de Isw1 y Chd1 se ha descrito que causa un incremento importante de la transcripción críptica a escala genómica (Hennig *et al.*, 2012).

Por otra parte, existen toda una serie de factores que modifican post-traduccionamente las histonas, remodelando así la cromatina al paso de la RNAPII durante las fases de inicio, elongación y terminación de la transcripción. Sobre algunos de ellos se hablará en el apartado correspondiente.

2.1.5 Fases del ciclo transcripcional.

El proceso de transcripción por la RNAPII se puede dividir en tres fases: 1) la iniciación, en la cual la polimerasa se une al promotor del gen y recluta los factores necesarios para empezar a transcribir; 2) la elongación, en la que esta avanza sobre el molde de DNA sintetizando la molécula de RNA, y 3) la terminación y reinicio, en la que se finaliza la síntesis de la molécula de RNA naciente. Tras ello, la RNAP se libera del DNA y, o bien vuelve a empezar otro ciclo, o bien es degradada (Mischo y Proudfoot, 2012).

2.1.5.1 Iniciación

Como se ha comentado anteriormente, uno de los primeros pasos de la iniciación de la transcripción es la fosforilación de la Ser5 del CTD por la quinasa la subunidad del Mediador Srb10. La eliminación de la marca del CTD promovida por Srb10 es un paso clave en el proceso de iniciación, ya que hasta que ésta no es eliminada, la RNAPII no es capaz de ensamblarse funcionalmente al PIC. La transcripción se inicia con la formación del complejo de promotor cerrado (*closed promoter complex*), que es el resultado del

contacto directo de la RNAPII con el DNA, en el contexto de un PIC completamente ensamblado (Cheung y Cramer, 2012). Tras ello, Kin28 fosforila también la Ser5 (Komarnitski *et al.*, 2000). Esta marca de la RNAPII recluta toda una serie de factores que van a actuar en etapas posteriores, incluyendo a la enzima que incorpora la caperuza al RNA naciente, el sistema de vigilancia de la integridad del mRNA Nrd1-Nab3, y el complejo PAF, que conecta la fosforilación en Ser5 con toda la red de modificaciones de histonas posteriores. Existen evidencias de que la RNAPII escanea la secuencia del DNA del promotor para encontrar un TSS apropiado. El mecanismo propuesto para ello requiere el paso de un complejo cerrado a uno abierto, en el que el DNA alrededor del TSS se desenrolla por mediación de la helicasa dependiente de hidrólisis de ATP del factor TFIIH (Rad25) (Hahn y Young, 2011), formándose la burbuja de transcripción. Además de TFIIH, el escaneo del promotor depende de TFIIIB y de TFIIIF (Kostrewa *et al.*, 2009). El DNA de simple cadena se inserta en el sitio activo de la RNAPII y empieza la polimerización con los NTPs como sustratos. El híbrido DNA/RNA inicial no es lo suficientemente largo como para formar un complejo estable de elongación, con lo que se producen varias rondas de síntesis de RNAs nacientes cortos en lo que se conoce como la fase de iniciación abortiva (Shandilya y Roberts, 2012). Cuando el RNA alcanza una longitud mínima de 6 nt los factores de iniciación se liberan, se pierden contactos con el promotor, y se forma un complejo estable de elongación que escapa del promotor (Zawel *et al.*, 1995; Cheung y Cramer, 2012). Una vez la RNAP comienza la elongación, algunos GTFs entre ellos TFIID, A y B, así como el Mediador permanecen en el promotor (Pokholok *et al.*, 2002), lo que permite la rápida entrada de nuevas RNAPs para reiniciar la transcripción. Por el contrario, otros factores, como el complejo de chaperonas de histonas FACT, Spt4, Spt6, Paf1, el complejo involucrado en la exportación de mRNAs al citoplasma TREX, el co-activador SAGA, y los remodeladores de cromatina del grupo SWI/SNF viajan con la RNAPII a través de la zona codificante (Sims *et al.*, 2004).

2.1.5.2 Elongación

La elongación empieza con la salida del complejo de elongación (CE) del promotor y la síntesis procesiva del RNA naciente. Cuando la RNAP encuentra lesiones en el DNA, o cuando incorpora nucleótidos de forma errónea, puede moverse aguas arriba sobre sus

pasos (*backtracking*) y entrar en parada. Algunos factores negativos pueden frenar la elongación, o inducir pausas en esta región, produciendo un mecanismo de regulación por pausas próximas al promotor. Este mecanismo, en el que está involucrado los complejos DSIF y NELF, se ha demostrado importante en eucariotas superiores (Fromm *et al.*, 2013; Fuda y Lis, 2013), pero menos en levadura (Mayer *et al.*, 2010). Una vez pausada, la RNAPII necesita de factores adicionales para reanudar su actividad, entre los que se encuentran TFIIIS, que estimula el corte del RNA naciente deslocalizado del sitio activo durante la pausa (Gómez-Herreros *et al.*, 2012), o el factor Spt4/5, que estabiliza el CE, aumentando la procesividad del holoenzima (Martinez-Rucobo *et al.*, 2011). Adicionalmente, se ha descrito que uno de los factores principales de degradación citoplasmática de los mRNAs, el complejo Ccr4-NOT, también actúa como factor de elongación transcripcional, ayudando a las RNAPs paradas a reanudar la síntesis de transcritos (Kruk *et al.*, 2011).

Una vez en conformación estable, y salvo pausas o bloqueos, la síntesis pasa a la fase de elongación productiva. Para que eso sea así, la correcta modificación de la cromatina que cubre el gen es esencial para permitir el paso del CE. De ello se encargan toda una serie de factores modificadores de histonas, entre los que cabe destacar los complejos **Elongador**, **COMPASS** y **FACT**, la acetiltransferasa **Gcn5**, la metiltransferasa **Set2**, y la carabina de histonas **Spt6**.

El **Elongador** es un complejo multiproteico altamente conservado en eucariotas, formado por dos subcomplejos, que se unen a la forma fosforilada de la RNAPII y cuya principal función es la acetilación de histonas (Krogan y Greenblatt, 2001).

El complejo **COMPASS** está formado por siete proteínas, de las cuales la principal es Set1. Esta proteína posee actividad histona metiltransferasa (HMT) y depende del complejo de ubiquitinación de histonas Paf1 para su unión a la RNAPII.

El complejo **FACT**, compuesto por Spt16 y Pob3, posee un papel importante en el reensamblaje de la cromatina tras el paso de la RNAP (evitando así la transcripción críptica) y también de la DNA Polimerasa durante la replicación (Whittmeyer y Formosa, 1997). Existen evidencias que indican que la transcripción por la RNAPII es especialmente dependiente de FACT en unidades transcripcionales que presentan una cromatina bien

posicionada en la región 5' cercana al promotor, como los genes RPs (Jimeno-Gonzalez *et al.*, 2006; Pelechano *et al.*, 2009, Douglas Maya, tesis doctoral, 2012).

La acetiltransferasa **Gcn5**, componente del complejo activador SAGA, fue la primera proteína aislada con actividad HAT, y tiene un papel en la remodelación de cromatina en la fase de elongación (Govind *et al.*, 2007; Kim *et al.*, 2010). En asociación estrecha con SAGA se encuentra otro complejo con funciones HAT que regula la expresión de un grupo pequeño de genes, el complejo SLIK (SAGA-Like) (Wu y Winston, 2002).

Set2 es otra proteína con actividad HMT (Strahl *et al.*, 2002), que interacciona una vez la RNAPII se empieza a fosforilar en la Ser2 del CTD. La metilación de histonas por parte de Set2 se asocia con la región codificante de genes activos.

Spt6 está conservado en todos los eucariotas, y, de manera similar a Spt16, es necesario para reensamblar los nucleosomas tras el paso de la RNAP, evitando la aparición de transcripción críptica (Kaplan *et al.*, 2003). Además, hay evidencias que lo posicionan como un elemento de coordinación entre la elongación de la transcripción y el metabolismo postranscripcional del mRNA. En este sentido, se ha visto que se asocia al exosoma nuclear, un complejo exoribonucleasa encargado del control de calidad de los mRNAs (Andrulis *et al.*, 2002), que existe una relación funcional con factores de relacionados con la poliadenilación en 3' de algunos genes (Bucheli y Buratowski, 2005), y que se recluta al complejo de exportación TREX (Yoh *et al.*, 2007).

2.1.5.3 Terminación y reiniciación

A pesar de tratarse de un proceso fundamental de la expresión génica, la terminación es la etapa menos conocida del ciclo transcripcional (Hennig y Fischer, 2013; Anamika *et al.*, 2013). La terminación de la transcripción de los genes la RNAPII requiere dos procesos interconectados pero distintos: la formación de un extremo 3' definido, y la disociación de la RNAPII del molde de DNA. Cuando la RNAPII se acerca al final del transcrito, diversas señales de terminación hacen que se disparen los mecanismos de maduración del extremo 3' y terminación de la transcripción en sí. La formación de un

extremo 3' resistente a la degradación por nucleasas es llevado a cabo por la maquinaria de corte y poliadenilación (C/P). El reclutamiento de la maquinaria de C/P depende de varios factores, entre los que se encuentran el reconocimiento de elementos de secuencia específicos en el transcrito nascente denominados señales poliA (pA), y del estado de fosforilación del dominio CTD (Mayer *et al.*, 2012). Por otra parte, y según el modelo aceptado actualmente, existen al menos dos mecanismos que contribuyen a la terminación: el primero se basa en un cambio conformacional del complejo de elongación causado por la disociación de factores de elongación, seguida por la asociación de factores de terminación a medida que la RNAPII transcribe la región con señales de pA (denominada “transición 3’ en dos pasos”, Mayer *et al.*, 2010).

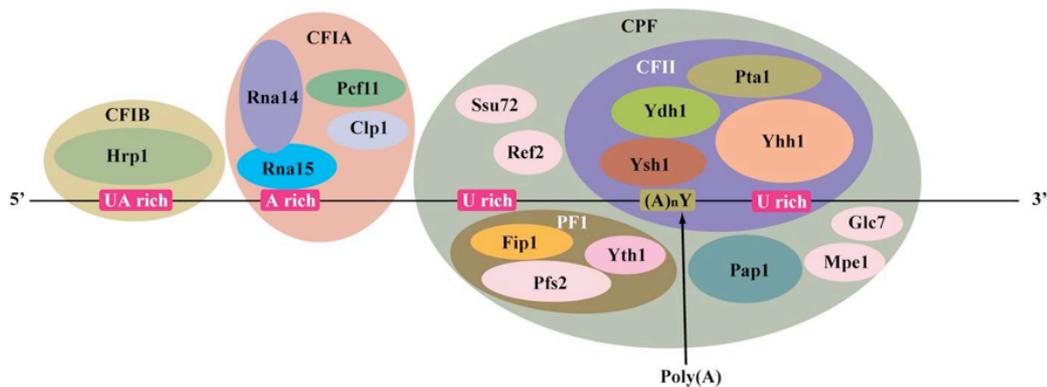


Figura i3. Señales de procesamiento del extremo 3' del pre-mRNA y complejos de C/P. Tomado de Yang y Doublé (2011).

El segundo involucra al denominado efecto “torpedo”, por el cual la rápida degradación del transcrito nascente asociado a la RNAPII más allá del sitio pA, mediada por la exonucleasa 5'-3' Rat1 (Xrn2), desplaza a la RNAPII del DNA molde (Richard y Manley, 2009).

Las tres RNAPs eucarióticas tiene la habilidad de cortar el RNA nascente en el interior de sus sitios activos. No obstante, estudios *in vitro* muestran que la RNAPII requiere de toda una serie de complejos proteicos para la fase de corte en 3'. La mayoría de los transcritos de RNAPII se poliadenilan en 3', pero muchos otros no son poliadenilados, con lo que su estabilización depende de interacciones específicas RNA-proteína.

-Procesado y terminación de transcritos codificantes de proteínas: en *S. cerevisiae*, cuando la RNAPII transcribe el sitio pA, el RNA naciente es cortado endonucleolíticamente y protegido de la degradación por 3' mediante la adición de un homopolímero de 70 adeninas, la cola pA (Bienroth *et al.*, 1993). El responsable de esta tarea es el "poliAsoma", un macrocomplejo de unas 20 proteínas organizadas en varios subcomplejos (Proudfoot, 2011). Al contrario que en humanos, donde se encuentra la secuencia consenso de poliadenilación AAUAAA, en levadura las señales son generalmente degeneradas, difíciles de reconocer. Para la mayoría de genes, consisten en un elemento de eficiencia EE rico en adeninas, un elemento posicionador PE también rico en adeninas y un elemento rico en uridinas que cubre tanto el sitio de corte como el de adición de la cola pA (van Helden *et al.*, 2000). En el momento en que una señal pA es transcrita, los complejos CPF (*Cleavage and Polyadenylation Factor*) y CF (*Cleavage Factors IA y IB*) se unen al transcrito naciente (Ansari y Hampsey, 2005). Algunos de los factores que componen estos complejos también se unen al dominio CTD durante la elongación. Es el caso por ejemplo de los factores Nrd1 y Pcf11, que se unen al CTD fosforilado en Ser2 (Lunde *et al.*, 2010). Según se ha mencionado en apartados anteriores, el CTD se fosforila en residuos distintos conforme avanza la transcripción, a modo de mecanismo de coordinación. A nivel de terminación, se ha descrito que las fosforilaciones en Ser5, y especialmente en Tyr1, son inhibitorias de la unión de los factores de terminación (Mayer *et al.*, 2012). Una vez unidos al pre-mRNA, los complejos CPF/CF catalizan el corte endonucleolítico y la adición de la cola pA. Concretamente, Ysh1 se encarga del corte (Ryan *et al.*, 2004), y la poliA polimerasa Pap1 de la adición de la secuencia de poliadenilato (Ezeokonkwo *et al.*, 2012). El carácter degenerado de las señales pA en levadura resulta en cambios en la selección del extremo 3'. Recientemente, estudios genómicos han desvelado que muchos transcritos poseen sitios de C/P alternativos, lo que genera isoformas con diferentes potenciales codificantes y/o regiones UTR 3' variables (Pelechano *et al.*, 2013; Moqtaderi *et al.*, 2013). La cola pA protege de la degradación, y dicta la estabilidad del transcrito y su comportamiento en traducción. En estos procesos participan una serie de proteínas de unión a cola pA, PABPs (*PolyA Binding Proteins*), las cuales se encargan de mantener la adecuada extensión de la cola pA, ayudar en la exportación del transcrito al citoplasma, y modular su traducción. Las PABPs principales en *S. cerevisiae*

son Nab2 y Pab1. Recientemente se ha descrito que el componente del exosoma nuclear Rrp6 también actúa como PABP, controlando la longitud de la cola pA y modulando la unión de Nab2 (Schmid *et al.*, 2012).

A diferencia de la poliadenilación, la terminación de la transcripción se produce unos 200 nt aguas abajo del sitio pA en *S. cerevisiae*, y unos 1500 nt en humanos (Grosso *et al.*, 2012). Los mecanismos moleculares por los que la RNAPII se suelta del DNA molde no se conocen completamente, pero existen cada vez más evidencias de que esta disociación se basa en una disminución en la estabilidad del híbrido DNA:RNA, resultado de la acción combinada de muchos factores. Entre ellos, los más importantes serían la disminución en la velocidad del complejo de elongación, y la acción a modo de *torpedo* de la exonucleasa 5'-3' Rat1. La disminución de la tasa de elongación, que conlleva la pausa del complejo (Kazerouninia *et al.*, 2010), es resultado de la unión de los factores de terminación, lo cual sería aprovechado por Rat1 para desestabilizar el híbrido DNA:RNA y soltar a la RNAP. Otro factor adicional que ayuda a la terminación es la presencia de nucleosomas posicionados después del sitio pA. Al encontrarse con los nucleosomas, la RNAPII se pausa pero no puede sobrepasar el obstáculo, ya que carece de factores de elongación como TFIIIS/Dst1, los cuales se han disociado del complejo durante la primera fase de la transición 3' (Grosso *et al.*, 2012).

-Procesado y terminación de transcritos no codificantes: respecto a los transcritos no codificantes, se han descrito varios mecanismos de procesado y terminación, dependientes del tipo de ncRNA. Por ejemplo, los SUTs y XUTs se procesan de manera similar a los transcritos codificantes, mientras que otros ncRNAs, como los snoRNAs, o los CUTs, dependen de una combinación de componentes de los complejos CPF/CF y NRD (también llamado complejo Nrd1-Nab3-Sen1). Entre los componentes más importantes del complejo NRD encontramos a la helicasa Sen1, la cual actúa desapareando directamente el híbrido DNA:RNA del sitio activo de la RNAPII, provocando su disociación. Además de en ncRNAs, se ha descrito que Sen1 es capaz de actuar en genes codificantes, quizá facilitando la acción de Rat1 (Steinmetz *et al.*, 2006; Kawauchi *et al.*, 2008).

-Procesado y terminación de transcritos crípticos: cuando la RNAP se ensambla al DNA molde, la transcripción puede ser dirigida en ambas direcciones desde el PIC (Wei *et al.*, 2011). De esa manera, la célula necesita de un mecanismo que termine los transcritos que empiezan en la dirección no favorecida. Por otra parte, cuando no todos los factores requeridos para la elongación productiva se encuentran en el promotor, la RNAPII terminará de manera rápida. En ambos casos, se ha descrito que la vía de terminación de esos transcritos nacientes sería la dependiente de NRD (Creamer *et al.*, 2011). Otro caso es el que se da cuando la RNAP no consigue terminar en la zona oportuna, y continúa transcribiendo aguas abajo. Este fenómeno ocurre hasta en el 60% de los complejos de elongación en genes altamente transcritos en levadura (Rondon *et al.*, 2009), y para su terminación la célula hace uso de cualquiera de los mecanismos de los que dispone de manera oportunista (Mischo y Proudfoot, 2012).

-Reiniciación rápida de la transcripción y memoria transcripcional: cada vez hay más evidencias de la existencia de interconexiones funcionales entre factores de iniciación y de terminación en determinados grupos de genes. Dichos genes poseen un mecanismo de regulación adicional a los descritos hasta ahora, dependiente de la adopción de estructuras cromatínicas espaciales en forma de lazo. Estas estructuras, en las que las regiones promotoras y terminadoras están yuxtapuestas, permiten una rápida reiniciación de la transcripción por la misma RNAP que ha completado la ronda anterior (Kuehner *et al.*, 2011). Es lo que se conoce como *gene looping*. Aunque se descubrieron primero en *S. cerevisiae* (O'Sullivan *et al.*, 2004; Ansari y Hampsey, 2005), los lazos génicos ya han sido encontrados en ratones y humanos (Grosso *et al.*, 2012). Los lazos han sido detectados por técnicas como el 3C (*Chromosome Conformation Capture*) que mapean contactos físicos entre regiones del genoma, y requieren para su formación de la participación de componentes del PIC (TFIIB), de la maquinaria de procesado del extremo 3' (Ssu72, Sub1), y de subunidades del complejo del poro nuclear (Hampsey *et al.*, 2011). La formación del lazo parece ser preferente en los genes cortos (Grosso *et al.*, 2012) y ser dependiente de que los genes sean activamente transcritos, ya que mutaciones en la subunidad Rpb1 de la RNAPII afectan a su formación (Ansari y Hampsey, 2005). Además, se ha propuesto que la reiniciación rápida conseguida de esta manera sería importante para diversos grupos de genes en los que el proceso de reciclado tuviera que ser muy eficiente, como es el caso de los genes RP. Por otra parte,

los modelos actuales proponen que los lazos afectarían o regularían tanto la direccionalidad de la transcripción, como el fenómeno de memoria transcripcional. La direccionalidad hace referencia a que el comienzo de una ronda pionera en sentido favorece la transcripción frente a antisentido en las siguientes rondas desde promotores bidireccionales (Tan-Wong *et al.*, 2012). En cuanto a la memoria transcripcional, se propone que una primera ronda de transcripción pionera sería necesaria para formar un andamio de reiniciación rápida que quedaría preparado para cuando un gen tuviera que ser transcrito de nuevo rápidamente (Brickner, 2010; Hampsey *et al.*, 2011).

2.1.6 Aspectos asociados a la transcripción

2.1.6.1 Procesamiento de pre-mRNAs

El estado de fosforilación del dominio CTD no solamente marca la progresión a través del ciclo de transcripción, sino que también regula el reclutamiento co-transcripcional de los diferentes factores implicados en el procesamiento del mRNA: la adición de la caperuza, el *splicing*, y la poliadenilación del transcrito, entre otros. Aunque estos procesos ocurren secuencialmente, están fuertemente interconectados y coordinados por el dominio CTD, que actúa como plataforma de unión de los distintos complejos implicados en la maduración del transcrito.

-Adición de la caperuza: tras la síntesis de alrededor de 20 a 30 nt, el extremo 5'-trifosfato del pre-mRNA es modificado a través de la adición de una caperuza. Esta transformación consiste en la adición de un grupo 7-metilguanosina sobre el enlace trifosfato del primer nucleótido transcrito que previene la degradación 5'-3' de los RNAs nacientes. En la adición de la caperuza, la RNAPII se detiene próxima al promotor y la fosforilación de la Ser5 del dominio CTD proporciona un sitio de unión para las enzimas encargadas de dicha transformación (Coppola *et al.*, 1983; Komarnitsky *et al.*, 2000). Una vez completada la adición de la caperuza, la RNAPII continúa con la adición de nucleótidos. Esta parada y

reactivación de la RNAPII actúa como punto de control asegurando que sólo los transcritos que contienen la caperuza continúan la elongación transcripcional (Pei *et al.*, 2003).

-Eliminación de intrones: la proporción de genes con intrones en levadura es mucho menor que la que presentan otros eucariotas. Sólo el 3% de los genes de *S. cerevisiae* presentan intrones y de estos, el 98% contiene un único intrón pequeño de menos de 300 nt. Los intrones son eliminados por un complejo denominado espliceosoma formado por cinco snRNPs (U1, U2, U4 y U6) constituido por la asociación de un molécula de snRNA con diversas proteínas (Blencowe, 2006). Múltiples trabajos muestran que algunos componentes del espliceosoma se reclutan co-transcripcionalmente a través de su asociación con las formas fosforiladas en Ser5 y Ser2 del dominio CTD de la RNAPII (Gornemann *et al.*, 2005; Lacadie *et al.*, 2006; Tardiff *et al.*, 2006). El ensamblaje de un espliceosoma catalíticamente activo depende tanto del reclutamiento secuencial de los factores de corte y empalme, como de las modificaciones estructurales necesarias para la formación de dos centros catalíticos sobre el RNA naciente (Wahl *et al.*, 2009). Esta cascada de eventos de ensamblaje requiere una cantidad considerable de tiempo. Se estima que el tiempo necesario para el procesado de intrones en levadura es de aproximadamente 30 seg. El *splicing* ocurre co-transcripcionalmente si la catálisis termina antes que termine la transcripción. Por ello, se considera que el procesado de intrones co-transcripcional “compite cinéticamente” tanto con la elongación, como con el corte del RNA naciente en el sitio pA, por lo que siempre se ha creído que el *splicing* ocurría post-transcripcionalmente (Tardiff *et al.*, 2006). Esta creencia ha sido rebatida recientemente con la aparición de nuevos trabajos que demuestran que la mayoría de intrones se procesan co-transcripcionalmente tanto en levadura (Carrillo Oesterreich *et al.*, 2010; Alexander *et al.*, 2010) como en *Drosophila* (Khodor *et al.*, 2011). El mecanismo por el que esto es posible se ha denominado *Terminal Exon Pausing*, y en él la RNAPII se pausa al final de las zonas intrónicas (en el sitio de *splicing* 3') el tiempo necesario para permitir el reclutamiento del espliceosoma completo y su ensamblaje en el RNA naciente (Carrillo Oesterreich *et al.*, 2011). Recientemente se ha demostrado que la

velocidad de la RNAPII está inversamente relacionada con la eficiencia de *splicing*, lo que concuerda con este modelo (Braberg *et al.*, 2013).

2.1.6.2 Coordinación entre síntesis y degradación de mRNAs

El mRNA es un intermediario inestable encargado de la transmisión de la información genética desde el DNA hasta los ribosomas, donde es transformado en proteínas. Además de los mRNAs que siguen esta ruta, existen muchas otras clases de RNAs funcionales que no se traducen a proteínas. Entre ellos encontramos RNAs tan abundantes en la célula eucariota como los RNAs de transferencia (tRNAs) o los RNAs ribosómicos (rRNAs), y otros más minoritarios como los transcritos crípticos (SUTs, CUTs, XUTs.), snoRNAs, microRNAs, siRNAs, snRNAs, piRNAs, RNAs TERRA, o RNAs no codificantes largos (*long ncRNAs*). Estas clases de RNAs están involucradas en la regulación de un amplio abanico de procesos, como pueden ser el mantenimiento de la estructura cromosómica, la transcripción, el *splicing* y la traducción. Las macromoléculas tienen una vida útil limitada temporalmente en las células, y los RNAs, que además carecen de mecanismos de corrección de errores conocidos, no son una excepción. Al terminar su vida útil, o cuando los mecanismos de control de calidad encuentran fallos durante su síntesis, los RNAs son degradados. Así, se puede considerar que el contenido de mRNAs de una célula es el resultado tanto de su síntesis como de su degradación.

Todas las rutas de degradación de mRNAs culminan en la actuación de unas pocas nucleasas, ayudadas por multitud de factores accesorios. Las actividades de degradación 5' a 3' y 3' a 5' existen en todos los organismos, mientras que las actividades de eliminación de la caperuza y de la cola pA son exclusivas de los eucariotas. La degradación de mRNAs es un proceso altamente eficiente, siendo esa la causa de que algunos productos de la RNAPII sean difíciles de detectar, como los intrones procesados o muchos ncRNAs.

Los principales mecanismos de degradación de RNAs se pueden diferenciar según actúen en el núcleo o en el citoplasma:

-Rutas de degradación nucleares: La síntesis de pre-mRNAs está acoplada a la adición de una caperuza en 5', que confiere protección frente a exonucleasas-5', y de una cola de pA, que protege los extremos 3'. Desde el principio de su síntesis, los pre-mRNAs se asocian a proteínas, y posiblemente también con ncRNAs, para formar mensajeros ribonucleoprotéicos, mRNPs (*messenger RiboNucleoProteins*). Defectos en el procesado nuclear de mRNAs y de la formación de mRNPs capaces de ser exportados al citoplasma pueden promover su degradación por los mecanismos de vigilancia y control de calidad nucleares. El factor principal que ejecuta este mecanismo de degradación es el complejo exosoma nuclear, cuya subunidad central es la exonucleasa 3'-5' Rrp6. Rrp6 lleva a cabo un control de calidad de la formación del extremo 3', y su mutación puede resultar en terminación prematura y defectos en la poliadenilación (Chlebowski *et al.*, 2013). La actividad del exosoma nuclear está asistida por el complejo TRAMP. Este complejo consiste en dos proteínas de unión a RNA, Air1 y Air2, una helicasa de RNA, Mtr4, y una RNAP no convencional, Trf4p (LaCava *et al.*, 2005).

Otro mecanismo de vigilancia detecta pre-mRNAs cuyos intrones no han sido procesados, y los degrada mediante el exosoma, o de 5'-3' mediante la exonucleasa Rat1 (Bousquet-Antonelli *et al.*, 2000). Por su parte, los transcritos no *tradicionales* que se producen en el núcleo (intergénicos, intragénicos, asociados a promotores, a regiones heterocromáticas y antisentidos) también son diana de factores de degradación nucleares como el exosoma (Milligan *et al.* 2008), en asociación con TRAMP (Callahan y Butler, 2010). El complejo TRAMP puede promover la degradación 3'-5' de esos transcritos añadiéndoles una cola pA, lo cual recluta al exosoma (Wyers *et al.*, 2005). Además de esos transcritos, también es capaz de controlar la degradación de otros tipos de RNAs como los snoRNAs, snRNAs, o tRNAs. A su vez, el complejo TRAMP está modulado en parte por las proteínas de unión a RNA Nrd1 y Nab3, las cuales se unen a elementos específicos del RNA naciente (Cramer *et al.*, 2011; Kim y Levin., 2011) y al dominio CTD cuando éste es fosforilado en Ser5 y Ser7 (Kim *et al.*, 2010), con lo que podría estar actuando preferentemente sobre transcritos cortos, dado que tales

fosforilaciones se van perdiendo conforme avanza la elongación (Komarnitski *et al.*, 2000).

Por tanto, la degradación nuclear se encarga, por una parte, de eliminar mRNAs defectuosos, y por otra parte, de degradar algunos transcritos no convencionales. La mayoría de moléculas funcionales, y algunas erróneas, escapan de estos mecanismos de control y alcanzan el citoplasma, donde son supervisadas de nuevo por otros mecanismos, detallados a continuación.

-Rutas de degradación citoplasmáticas:

-Degradación de mRNAs defectuosos: los mRNAs citoplasmáticos cuyos defectos les impiden ser traducidos eficientemente son diana de mecanismos de vigilancia cuyo objetivo es asegurarse que los recursos energéticos y de la maquinaria traduccional no se emplean en la producción de proteínas aberrantes. El mecanismo mejor conocido es el de degradación de mRNAs sin sentido, NMD (*Non-cense-Mediated Decay*) (revisado en Shoenberg y Maquat, 2012). El NMD degrada los mRNAs cuya traducción acaba prematuramente, como consecuencia por ejemplo de la presencia de codones de parada prematuros, o de inicios de traducción alternativos que alteran la pauta de lectura, entre otros. Es un mecanismo conservado en todos los eucariotas, en el que están involucrados los factores Upf1, Upf2, Upf3, y la exonucleasa Xrn1. La exonucleasa 5'-3' Xrn1 es el componente principal de degradación de mRNAs en el citoplasma, actuando tanto sobre mRNAs maduros como sobre otras especies de mRNAs defectuosas que escapan a los mecanismos de control nucleares. Existen otros dos mecanismos alternativos al NMD en eucariotas: los denominados *no-go decay* y *non-stop decay*.

-Degradación de mRNAs funcionales: la degradación de los mRNAs maduros ocurre mayoritariamente en el citoplasma, mediante diferentes mecanismos (Parker, 2012). El primer paso suele ser el acortamiento o la completa

eliminación de la cola pA, llevada a cabo por tres complejos de desadenilación: Pan2/Pan3, Ccr4-NOT y PARN (Goldstrohm y Wickens, 2008). El mRNA desadenilado puede ser degradado exonucleolíticamente de 3'-5' por el exosoma citoplasmático (Anderson y Parker, 1998). El exosoma citoplasmático es un complejo multiproteico compuesto de 9 subunidades principales y otras accesorias (Durand y Lykke-Andersen, 2011). Además, es común que el acortamiento de la cola pA vaya acompañado de la eliminación de la caperuza en 5' por el complejo Dcp1/Dcp2, y la subsiguiente degradación 3'-5' mediada por Xrn1 (Muhlrad *et al.*, 1994).

Todos los enzimas y factores asociados involucrados en los mecanismos generales de degradación de mRNAs son putativos puntos de control de la expresión génica. En *S. cerevisiae*, las mutaciones en muchos de estos factores muestran fenotipos pleiotrópicos, lo que indica que los factores de degradación podrían tener un papel en la coordinación de la expresión génica, conectando la degradación con otros procesos como la traducción, la transcripción, o la exportación nuclear (Pérez-Ortín *et al.*, 2013). A este respecto, existen evidencias experimentales que muestran que una mutación de la RNAPII que compromete la transcripción no afecta en la misma proporción a la cantidad de mRNAs maduros (Goler-Baron *et al.*, 2008, Sun *et al.*, 2012). De manera similar, la eliminación del complejo Ccr4-NOT afecta la tasa de degradación pero no afecta de la manera esperada a la tasa de transcripción (Sun *et al.*, 2012). A pesar de que los mecanismos siguen siendo desconocidos, estos resultados pueden ser explicados globalmente por la existencia de un mecanismo compensatorio de las tasas de transcripción cuando la degradación se compromete, y viceversa. Estudios recientes confirman que el diálogo cruzado entre los mecanismos de transcripción y degradación es más amplio de lo que se creía, y que están involucrados factores tan importantes como el submódulo de la RNAPII Rpb4/Rpb7, el TF Rap1, y los factores de degradación Ccr4-NOT y Xrn1, que viajan del núcleo al citoplasma y viceversa para controlar y modular los procesos de transcripción y degradación (Dahan y Choder, 2013, Haimovich *et al.*, 2013). Estos factores forman parte de una nueva categoría de elementos celulares denominados “sintegradasas”, debido a su capacidad dual para estimular/reprimir tanto la síntesis como la degradación de los mRNAs.

3. Transcripción naciente

Muchas de las características distintivas de la expresión génica sólo pueden describirse de manera clara al comparar entre sí un número estadísticamente relevante de genes a la vez. Habitualmente, los experimentos de evaluación de las respuestas transcripcionales de las células a cambios ambientales o de otro tipo se han centrado en el estudio, en genes puntuales o a nivel genómico (transcriptómico), del cambio en cantidad de RNA total presente en la célula, RA (RNA *Amount*). Medir la RA directamente implica no tener en consideración que el RA es el resultado de los cambios coordinados de los procesos de síntesis y degradación de los mRNAs. Para entender mejor los mecanismos por los que una célula establece un RA determinado, y lo adapta a las perturbaciones o a los cambios inherentes a su ciclo celular, es interesante estudiar los procesos que operan a ambos lados de la balanza de la expresión génica: la síntesis de los transcritos, y su degradación. Hoy en día esto es posible gracias al uso de diversas técnicas genómicas y transcriptómicas que miden tanto la tasa de transcripción, TR (*Transcription Rate*), como la tasa de degradación, DR (*Degradation Rate*). Se entiende por TR la tasa a la que aparecen en el citoplasma las moléculas maduras de mRNA. La TR puede ser medida en moléculas por minuto u hora. Dado que existen muchos procesos de regulación desde que las RNAPs empiezan a transcribir hasta que un transcrito maduro aparece en el citoplasma, la medida de TR al final de ese proceso, conocida como SR o tasa de síntesis de mRNA (*Synthesis Rate* SR, Pérez-Ortín *et al.*, 2013), no nos informa de qué está sucediendo en etapas anteriores del proceso, muy importantes para entender al detalle las dinámicas de la expresión génica. Para indagar en el proceso de la transcripción dentro del núcleo a nivel genómico se han desarrollado técnicas que miden la tasa a la que una molécula de RNA es sintetizada sobre su molde, el DNA. A esa TR se la llama TR naciente (en adelante nTR, *nascent Transcription Rate*, Pérez-Ortín *et al.*, 2013). Muchos de los métodos que se han desarrollado para la medida de TRs se basan en realidad en la medición de nTRs, debido su la mayor facilidad de implementación.

3.1 Técnicas genómicas para la medición de tasas de transcripción

Existen diferentes técnicas para medir la TR a nivel genómico:

-GRO y sus variantes: el GRO (*Genomic Run-On*) es una técnica por la cual los transcritos nacientes son marcados *in vivo* con nucleótidos radiactivos (García-Martínez *et al.*, 2004). Los transcritos marcados se hibridan en macrochips de nailon que contienen sondas de la zona de la ORF de todos los genes de *S. cerevisiae*. La señal correspondiente a una sonda determinada del chip es un indicador de la densidad promedio de RNAP. Si se asume velocidad de elongación constante, entonces esa densidad de RNAP es directamente proporcional a la TR. El GRO ha posibilitado la obtención del primer conjunto de datos robusto de nTR para un organismo (Pelechano *et al.*, 2010). Entre las variantes que se han desarrollado a partir de él encontramos el GRO-seq (Core y Lis, 2008), el NRO-seq (McKinlay *et al.*, 2011), el GRO-cap (Kruesi *et al.*, 2013) y el PRO-seq (Kwak *et al.*, 2013). Todas ellas parten del fundamento del GRO, pero difieren en el nucleótido que utilizan para hacer el marcaje, y en que secuencian masivamente los RNAs nacientes en vez de hibridarlos en chips. Una característica importante de estas metodologías es que obtienen información de un subconjunto de RNAPs concreto: las RNAPs elongantes. Esto es así debido a que el marcaje *in vivo* sólo serán capaces de realizarlo aquellas RNAPs que sean competentes para la elongación, no aquellas que se encuentren ensambladas en el promotor pero no hayan empezado a transcribir, o las que se encuentren en estado de pausa o parada a lo largo del gen (*backtracked*). Otra ventaja importante de algunas de ellas es que permiten el posicionamiento de las RNAPs sobre el genoma con una resolución máxima (1 pb), y con especificidad de hebra. Entre las limitaciones, cabe destacar que el GRO radiactivo calcula un único valor promedio de nTR a lo largo del gen, y que no es específico de hebra. Por su parte, una de las desventajas de los métodos no radiactivos es que se basan en purificaciones y amplificaciones de la muestra que pueden introducir sesgos biológicos (detectar RNAs maduros en vez de nacientes).

-Aislamiento de transcritos nacientes asociados a la cromatina: existen maneras de capturar los transcritos nacientes que no dependen de marcajes *in vivo*. Hasta el

momento se han desarrollado dos, y ambas se basan en el aislamiento de los RNAs nacientes asociados a la propia RNAP elongante, sin necesidad de entrecruzar esta al DNA. Una técnica lo consigue mediante fraccionamiento cromatínico (Carrillo-Oesterreich *et al.*, 2010; Khodor *et al.*, 2011) y otra, el NET-seq, mediante fraccionamiento cromatínico asociado a IP de RNAPII (Churchman y Weissman, 2011). A diferencia del GRO y sus derivados, estas técnicas capturan los RNAs nacientes asociados a RNAPs, independientemente del estado de competencia elongante de éstas.

-ChIP: la técnica de IP de RNAPII adaptada a chips o a ultrasecuenciación es un método habitual para la medición de nTRs (Mayer *et al.*, 2010; Kim *et al.*, 2010). Las señales de hibridación (o las lecturas de secuencias) permiten generar mapas precisos de posicionamiento de RNAPs a lo largo del genoma. A partir de ellos, y bajo la misma asunción de tasa de elongación constante del GRO, se derivan las nTRs. Esta técnica tiene la ventaja de que se puede obtener información de los distintos estados transcripcionales de las RNAPs. Esto se consigue utilizando anticuerpos diferentes para las diferentes formas de la RNAP: total (subunidades concretas del holoenzima), en fase de elongación temprana (CTD fosforilado en Ser5), o en elongación productiva (CTD fosforilado en Ser2). Entre los inconvenientes encontramos principalmente dos: 1) habitualmente, cuando se trata de IP de RNAP total, se transforma en nTR la señal derivada de la presencia de RNAPs sobre el DNA que no tienen un transcrito naciente asociado, y que por tanto no están transcribiendo. 2) el método es inespecífico de hebra.

-Marcaje metabólico: otra posibilidad de estimar la TR, no basada en el estudio de la presencia/actividad de RNAPs, es medir la aparición de mRNAs marcados en el citoplasma durante un periodo determinado de tiempo. Esto se consigue introduciendo en los cultivos de células un precursor nucleotídico marcado, como por ejemplo el tiouracilo (Cleary *et al.*, 2005; Sun *et al.*, 2012) o la tiouridina (Kenzelmann *et al.*, 2007; Miller *et al.*, 2011). El RNA tiolado es posteriormente aislado, purificado e hibridado en microchips. Este método tiene la ventaja de que puede aplicarse a células vivas intactas, y el inconveniente de que la TR medida no es instantánea, sino que refleja todos los eventos

que ocurren durante el periodo de tiempo que dura el marcaje metabólico. A diferencia de la transcripción naciente nuclear, a este tipo de TR se la llama madura (SR), ya que se basa en la aparición de mRNAs totalmente procesados en el citoplasma.

-Cálculos indirectos: finalmente, la TR puede calcularse indirectamente si se asume que existe un estado estacionario de la expresión génica por el que las tasas de síntesis y degradación son iguales, y se conocen los parámetros de RA y DR. De esta manera:

$$SR = DR = Kd \times RA$$

Esta aproximación tiene la ventaja de que no precisa de ningún experimento, pero la desventaja de depender de lo precisos que sean los conjuntos de datos de RA y DR para no introducir sesgos y errores durante el procesado matemático de los datos. A pesar de ello, las estimas indirectas de TR son utilizadas de manera habitual en la literatura (Pelechano y Pérez-Ortín, 2010; Miller *et al.*, 2011; Sun *et al.*, 2012).

3.2 Diferencias entre TR naciente y síntesis de mRNAs maduros

A pesar de que las distintas técnicas mencionadas en el apartado anterior estudian el mismo proceso, las diferentes aproximaciones experimentales puede generar resultados ligeramente distintos. Esto se debe en parte a los sesgos técnicos de cada metodología, y en parte también a diferencias biológicas resultantes de medir diferentes fases del ciclo transcripcional.

El término “tasa de transcripción” involucra dos conceptos relacionados, pero diferentes: la tasa de transcripción naciente (nTR), que mide la actividad transcripcional de las RNAPs *in situ* en los genes, y la tasa de síntesis de mRNAs maduros (SR), que mide la contribución de la transcripción a la concentración de los mRNAs celulares (Figura i4). Ambos parámetros son útiles para los biólogos moleculares, pero ni son intercambiables ni se expresan en las mismas unidades (revisado en Pérez-Ortín *et al.*, en prensa). Estos parámetros serían proporcionales si el volumen celular, el contenido de DNA de las células y la eficiencia de maduración de los mRNAs fueran constantes. Esta es una

situación fisiológica común, y que suele asumirse para comparar la nTR con la SR, pero hay que tener presente que no es una situación universal, por lo que lo más correcto sería utilizar la nTR para hablar del proceso transcripcional en sí (densidad de RNAPs sobre el DNA), y la SR para referirse a la contribución de la transcripción al balance de concentración de los mRNAs citoplasmáticos maduros que conduce a variaciones en la expresión génica.

Hechas estas consideraciones, lo que se sabe hasta el momento arroja la conclusión de que la nTR es la mejor manera de detectar genes regulados por un mismo programa transcripcional (regulones transcripcionales). Se ha comprobado que los cambios en nTR predicen mejor las relaciones funcionales entre genes que los cambios en RA, o en SR (Hayles *et al.*, 2010). Esto puede ser debido a que para su computo no se tienen en cuenta las etapas post-transcripcionales que pudieran estar afectando a la aparición de transcritos en el citoplasma, como la degradación nuclear o diferencias en la exportación. La nTR también parece ser la mejor manera de clasificar los genes de acuerdo a sus marcas en la cromatina. Dado que la nTR mide la tasa de elongación, afectada en gran medida por la organización de la cromatina, su comparación con conjuntos de datos de marcas específicas de cromatina *favorable* para la transcripción correlaciona mejor que con la SR o la RA. Por ejemplo, se parecen más los datos de trimetilación en H3K36 o de IP de la HAT Gcn5 con la nTR que con la RA o la TR indirecta (Pelechano *et al.*, 2010). La SR, no obstante, es más apropiada para estudios cinéticos, debido a su mayor relación con el estado estacionario del mRNA en el citoplasma, o su no dependencia de asunciones como la tasa de elongación constante.

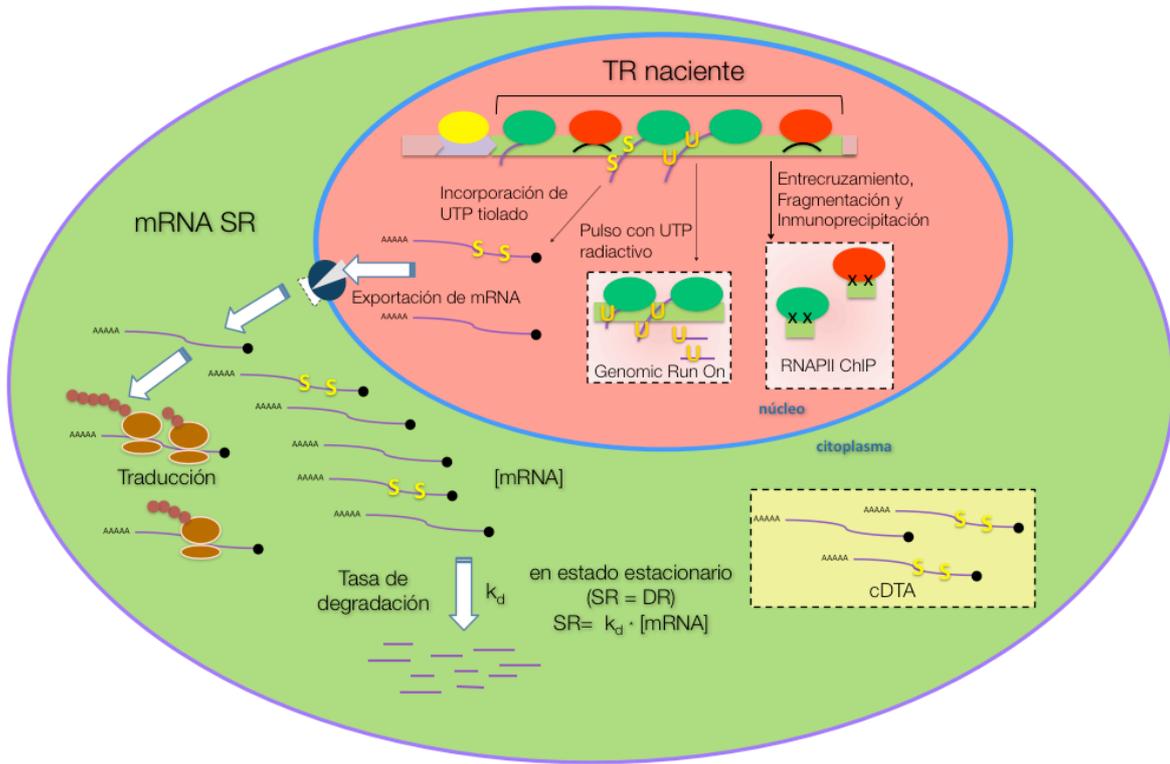


Figura i4. Esquema del proceso de la expresión génica, que muestra las diferentes técnicas de medición de nTR y SR. La nTR se determina a partir de la densidad de RNAP (óvalos coloreados) sobre moldes cromatínicos. Algunas moléculas de RNAP están en fase de inicio (amarillo), otras elongando (verde) y otras en estado de *backtrack* (rojo). Todas ellas pueden detectarse mediante ChIP, pero únicamente las elongantes darán señal de *run-on*. Si a las células se les da un pulso con precursores de UTP tiolados, el RNA recién sintetizado se puede purificar y cuantificar junto con el mRNA total no marcado. Las técnicas actuales no son capaces de distinguir entre el mRNA que está de manera transitoria en el núcleo (supuestamente en pequeña proporción), del citoplasmático (mayoritario). Asumiendo condiciones de estado estacionario, el protocolo de cDTA permite la determinación de la SR. Adaptado de Pérez-Ortín *et al.* (en prensa).

La comparación entre los distintos tipos de determinación de la TR muestra una buena correlación en general (Pérez-Ortín *et al.*, 2011). Los datos de nTR correlacionan mejor entre ellos que con los de SR, y los de SR tienen mejor correlación entre ellos y con la RA que con la nTR. Estos resultados son lógicos, puesto que los transcritos nacientes están sujetos a procesos post-transcripcionales de regulación hasta que aparecen en forma madura en el citoplasma.

Por tanto, se puede decir que cada conjunto de datos de TR tiene sus ventajas e inconvenientes, y que cada uno es más o menos apropiado según el objetivo del estudio. Así, de la comparación de las diferentes metodologías es de donde se extrae la información más amplia y valiosa sobre los mecanismos de regulación de la expresión génica.

3.3 Perfiles de densidad de RNAP como reflejo de diferencias en la tasa de elongación

Las medidas de nTR obtenidas por algunas de las técnicas mencionadas anteriormente (GRO y derivados, CHIP, NET-Seq) no se basan en la medición directa del número de moléculas de mRNA producidas en un determinado espacio de tiempo, sino que son computadas a partir del número de moléculas de RNAPII que hay en los genes.

La densidad de RNAPII sobre los genes depende mayoritariamente de tres parámetros: la velocidad de elongación (bases añadidas por unidad de tiempo), la frecuencia de inicios (número de eventos de iniciación que resultan en elongación productiva por unidad de tiempo) y de la procesividad (el número de RNAPs que permanecen en el DNA después de cada translocación) (Ehrensberger *et al.*, 2013). Para transformar densidad de RNAP sobre un gen en TR es necesario asumir que la velocidad de elongación es constante a lo largo de toda la región, y que además no hay transcripción abortiva por *drop-off* de RNAPs. Dichas asunciones pueden resultar útiles pero también simplistas, ya que hay muchas evidencias de que la elongación de la transcripción es un proceso altamente regulado. Un ejemplo reciente es el hallazgo *in vivo* de que la velocidad de elongación varía tanto con la temperatura de crecimiento como con la concentración de NTPs (Miguel *et al.*, 2013). Bajo estas asunciones, las nTRs calculadas no serían las reales, serían estimadas o aparentes. Una manera de testar esta hipótesis es estudiando la densidad de las RNAPs a lo largo de los genes, y tomando datos de TRs de regiones concretas. Por ejemplo, podemos calcular la tasa de inicio (TI), entendida como el número de moléculas de RNAPII que entran en fase de elongación productiva (decenas de nucleótidos aguas abajo del

TSS). También podemos calcular la tasa de terminación (TT), referida al número de moléculas de RNAPII que llegan de forma productiva al sitio de pA, dando lugar a un transcrito primario libre poliadenilado. La TT, según esta definición, representa la nTR real.

La TI sería igual a la TT si no existieran “accidentes” de las RNAPs por el camino. Estos accidentes han sido descritos en diversos sistemas modelo. Las RNAPs sufren pausas que no siempre pueden ser reiniciadas o solventadas por mecanismos como el *backtrack* u otros (Daulny y Tansey, 2009). Por ello se puede asumir que hay una tasa de accidentes o *drop-off* (DO) que se puede definir como la diferencia entre TT y TI normalizada por unidad de longitud (L):

$$DO = \frac{TT - TI}{L}$$

Por tanto, dado que para el cálculo de los valores de nTR que se dan por gen se están promediando zonas con potenciales diferencias en la densidad de RNAPs, se está introduciendo sistemáticamente un sesgo en los datos dependiente de la longitud del transcrito que aleja la nTR aparente de la nTR real. Este sesgo podría estar afectando a las comparaciones entre las diferentes mediciones de nTRs y SRs.

4. La transcripción de la RNAPIII

En eucariotas, la tarea esencial de transcribir de manera precisa el genoma la llevan a cabo tres RNAPs nucleares multi-subunidad. Como hemos visto hasta este punto de la introducción, la RNAPII se dedica principalmente a la síntesis de RNAs codificantes de proteínas, y de la mayoría de los RNAs *no tradicionales*. Por su parte, las RNAPs I y III transcriben genes esenciales involucrados en procesos celulares fundamentales. Concretamente, la RNAPI es responsable de la transcripción de los rRNAs precursores, y la RNAPIII transcribe RNAs esenciales cortos, como el rRNA 5S, los tRNAs, o el RNA pequeño

U6, que forma parte de la maquinaria de *splicing*. La actividad de las RNAPI y RNAPIII está estrechamente co-regulada, y da cuenta de un 80% de la transcripción génica nuclear en una célula en división (Paule y White, 2000). A pesar de ello, la RNAPII ha polarizado la mayoría de la atención de los investigadores, dado su papel de transcripción de los genes codificantes de proteínas, con lo que se han conseguido describir detalladamente sus aspectos bioquímicos, estructurales y funcionales. En cambio, las otras RNAPs están peor caracterizadas, y su estructura molecular ha empezado a ser desvelada recientemente (Vannini, 2013).

4.1 Visión general de la estructura de la RNAPIII

La RNAPIII es un enzima que consta de 17 subunidades, 5 de las cuales son idénticas a las de RNAPII (Rpb5, Rpb6, Rpb8, Rpb10 y Rpb12). Otras 7 presentan homología estructural (C160, C128, AC40, AC19, C11, C17 Y C25, homólogas de Rpb1, Rpb2, Rpb3, Rpb11, Rpb9, Rpb4 y Rpb7, respectivamente) y 5 subunidades específicas (C37, C53, C82, C34 y C31). Ver tabla i1.

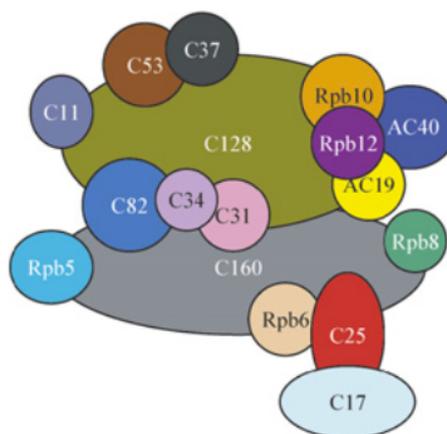


Figura i5. Modelo estructural del holoenzima de RNAPIII. Tomado de Lorenzen *et al.* (2007)

Las 5 subunidades específicas se organizan en dos subcomplejos funcionales: el heterodímero C37/C53, y el heterotrímero C82/C34/C31. Las estructuras cristalinas obtenidas y los datos de espectrometría de masas han revelado que la subunidad esencial C11, homóloga de la subunidad Rpb9 de RNAPII, se encuentra próxima al complejo C37/C53, y que posee una fuerte actividad de corte endonucleolítico 3' (Chedin *et al.*, 1998). Por su parte, el complejo C17/C25 parece ser estructural y funcionalmente parecido al submódulo Rpb4/7 de la RNAPII (Jasiak *et al.*, 2006).

Parte de la RNAP	Subunidad de RNAPIII	Subunidad de RNAPII	Tipo de subunidad
Core	C160	Rpb1	homóloga
	C128	Rpb2	homóloga
	AC40	Rpb3	homóloga
	AC19	Rpb11	homóloga
	Rpb5	Rpb5	común
	Rpb6	Rpb6	común
	Rpb8	Rpb8	común
	Rpb10	Rpb10	común
	Rpb12	Rpb12	común
	C11	Rpb9	homóloga
Heterodímero C17/C25	C17	Rpb4	contrapartida
	C25	Rpb7	contrapartida
Heterodímero C53/C37	C53	-	específica
	C37	-	específica
Hetrotrímero	C82	-	específica
	C34	-	específica
	C31	-	específica

Tabla i1. Relación de subunidades de la RNAPIII y sus homólogas con la RNAPII. *Homóloga* se refiere a que comparten homología funcional y de secuencia, mientras que *contrapartida* hace referencia a que sólo tienen homología funcional. Adaptado de Lorenzen *et al.* (2007).

La RNAPIII transcribe los siguientes tipos de genes en *S. cerevisiae*:

1. El rRNA 5S (locus RDN5): asociado a la subunidad larga del ribosoma.
2. snR6: RNA pequeño nuclear que forma parte del espliceosoma (U6 snRNA).

3. snR52: RNA pequeño nucleolar implicado en el procesado post-transcripcional del rRNA.
4. RPR1: componente de RNA de la ribonucleasa P, involucrada en la biogénesis de los tRNAs.
5. RNA170: RNA no codificante de función desconocida.
6. SCR1: RNA 7SL, componente de la partícula de reconocimiento de señal (SRP, *Signal Recognition Particle*), involucrada en la biogénesis de proteínas.
7. tRNAs: 274 genes de tRNA dispersos por el genoma, además de dos pseudogenes de tRNA (ZOD1 y iYGR033C).

4.2 Regulación de la transcripción de la RNAPIII

Las distintas condiciones de crecimiento de una levadura y los cambios fisiológicos que sufre afectan de manera sustancial a la transcripción de RNAPIII (Willis *et al.*, 2004). Entre ellos, cabe destacar diferentes fases del ciclo celular (como la entrada en fase G1 y la transición G2/M), la entrada en fase estacionaria, o el crecimiento bajo diversas condiciones de estrés (déficit de nutrientes, estrés oxidativo, térmico o por radiaciones ionizantes). A pesar de ello, los mecanismos moleculares y las rutas de señalización que controlan y modulan la expresión de los genes de RNAPIII sólo han empezado a conocerse en los últimos 10 años.

Estudios recientes han desvelado que el núcleo del PIC de las tres RNAPs eucarióticas está estructural- y funcionalmente conservado, y que es su periferia la que se decora con factores específicos de cada RNAP, necesarios para la selectividad específica de gen y su regulación (Vannini y Cramer, 2012). En general, el reclutamiento de cada RNAP a sus promotores específicos la llevan a cabo GTFs del promotor mínimo que contienen TBP y TFIIB (o una proteína parecida a TFIIB, la *TFIIB-like*). Los elementos del promotor que se posicionan aguas arriba y, en el caso de la RNAPIII, aguas abajo del TSS son reconocidos por otros complejos, específicos y no conservados, que ayudan al reclutamiento y posicionamiento preciso de los complejos que contienen a las proteínas TBP y *TFIIB-like*, irrespectivamente de la existencia de una caja TATA (Schramm y Hernandez, 2002).

La existencia de promotores con elementos internos localizados aguas abajo del TSS, dentro del cuerpo del gen, es una característica peculiar de la mayoría de los genes de RNAPIII, incluyendo el rRNA 5S y los tRNAs, controlados por lo que se conoce como promotores tipo I y tipo II, respectivamente (figura i6). En los promotores de tipo I, existe un elemento de control interno (ICR, *Internal Control Region*), reconocido por el TF TFIIIA, el cual recluta al complejo TFIIIC. Los promotores de tipo II poseen las llamadas cajas A y B, que son reconocidas directamente por el complejo TFIIIC. En ambos casos, el complejo TFIIIC recluta posteriormente al complejo TFIIIB, y este a su vez recluta a la RNAPIII. Existe otro tipo de promotor, el tipo III, que se encuentra únicamente en vertebrados. Este necesita de una caja TATA canónica y de otro elemento de secuencia con función reguladora llamado PSE, que son reconocidos por el complejo SNAPc. Una vez ensamblado SNAPc, se reclutan TFIIIB y después la RNAPIII (Teichmann *et al.*, 2010). TFIIIA es un TF que se une no sólo al promotor del gen 5S rRNA, sino también al transcrito naciente, modulando así su exportación al citoplasma (Pelham y Brown, 1980). El complejo TFIIIC consiste de 6 subunidades, y, aparte de su implicación en la transcripción de RNAPIII bajo promotores tipo I y II, se ha encontrado también fuera de las regiones transcritas por la RNAPIII, actuando como un barrera para el avance de la heterocromatina. Este hallazgo ha llevado a proponer a los sitios de unión de TFIIIC como potenciales marcadores globales para la cromatina (*chromatin bookmarks*) (Donze, 2012). El factor TFIIIB es un complejo trimérico compuesto por TBP, Bdp1 y Brf1, siendo el dominio C-terminal de este último el que sirve de andamio para la unión de la RNAPIII, a través de la subunidad C34 (Fernández-Tornero *et al.*, 2010).

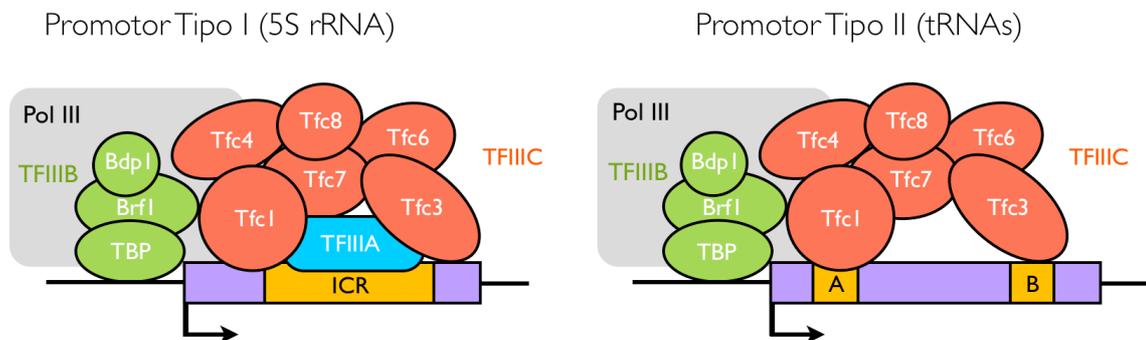


Figura i6. Tipos de promotores de RNAPIII en levadura. Adaptado de Vannini (2013).

Además de los elementos del promotor basal que se han tratado hasta este punto, existen otros reguladores de la actividad de la RNAPIII que merecen consideración:

Maf1: El único regulador negativo de RNAPIII conocido en levadura es la proteína Maf1, requerida para el apagado transcripcional que ocurre normalmente tras entrada en fase estacionaria, o en respuesta a diversas drogas, daño en el DNA, estrés oxidativo, defectos secretorios o crecimiento respiratorio (Desai *et al.*, 2005).

Sub1: caracterizada inicialmente como un coactivador de RNAPII, Sub1 se une al DNA e interacciona físicamente con activadores y componentes de la maquinaria basal de transcripción de la RNAPII. Además, está implicada en otras funciones celulares como la fosforilación del CTD durante la transcripción, el procesado del extremo 3' del mRNA y la terminación, la reparación del daño en el DNA y la organización de la cromatina (Conesa y Acker, 2010; García *et al.*, 2010). La primera evidencia de la participación de Sub1 en la transcripción de RNAPIII se obtuvo al co-precipitar Sub1 con TFIIC, y ver que Sub1 estimulaba la transcripción *in vitro* (Wang y Roeder, 1998). Posteriormente se ha descubierto que Sub1 se encuentra presente en el DNA de muchos genes de RNAPIII, incluyendo los tRNAs, facilitando el ensamblaje de TFIIB y TFIIC, estimulando así tanto la iniciación como la re-iniciación de la RNAPIII (Tavenet *et al.*, 2009).

TFIIS/Dst1: tal y como se menciona en el apartado 2.1.5.2, el factor TFIIS es conocido por ser una proteína de unión a RNAPII, a la que estimula fuertemente para producir el corte endonucleolítico de los RNAs nacientes. Su presencia en los genes de RNAPIII, detectada por CHIP, no era de esperar, ya que la propia subunidad C11 exhibe la misma fuerte capacidad de corte. Aún así, estudios genómicos han desvelado que la ausencia de TFIIS afecta la transcripción de la RNAPIII *in vivo* (Ghavi-Helm *et al.*, 2008).

Quinasas: algunas subunidades de la RNAPIII, TFIIC y Maf1 se fosforilan, y esta fosforilación es importante para la transcripción. Aproximaciones genéticas y bioquímicas han revelado que la proteína quinasa A (PKA), el complejo quinasa 1 TOR (TORC1), la quinasa regulada por TORC1 Sch9 y la caseína quinasa 2 (CK2) tienen un papel importante en la regulación de la transcripción de la RNAPIII (Acker *et al.*, 2013). CK2 por ejemplo es necesaria para la fosforilación de la TBP, y también para la liberación del represor Maf1

de los genes tRNA cuando se pasa de condiciones de crecimiento represivas a favorables (Graczyk *et al.*, 2011).

Remodeladores de cromatina y modificadores de histonas: estudios recientes a escala genómica han desvelado que los genes de tRNA están mayoritariamente deplecionados de nucleosomas, y que tienen una región 5' aguas arriba con nucleosomas bien posicionados (Mavrich *et al.*, 2008; Brogaard *et al.*, 2012). Entre los factores que mantienen esta estructura encontramos:

-Remodeladores de cromatina:

Nhp6: aunque no se conoce muy bien su mecanismo de acción, se sabe que Nhp6 se asocia *in vivo* a los genes tRNA y que está implicado en la iniciación de la transcripción, ayudando a mantener estructuras particulares de la cromatina en el promotor. También se sabe que interactúa con el complejo remodelador RSC (Kassavetis y Steiner, 2006), y con FACT (Formosa *et al.*, 2001).

FACT: este complejo tiene un papel central en la construcción, el mantenimiento, y el ajuste de la cromatina. Al igual que en los genes de RNAPII, actúa como chaperona de histonas, depositando una variante de la histona H2 (H2A.Z) en los nucleosomas de los genes de RNAPIII.

RSC: se ha descrito que RSC tiene un papel importante en la regulación de la cromatina aguas abajo del TTS de los tRNAs, y que esta regulación depende de la presencia de la RNAPIII sobre los genes (Kumar y Bhargava, 2013). También se ha descrito su papel en la regulación de los genes SNR6 y SUP4 (Arimbasseri y Bhargava, 2008).

Isw1 e Isw2: Isw1 se localiza en el cuerpo de los tRNAs, manteniéndolo libre de nucleosomas, e Isw2 ayuda a mantener la periodicidad del grupo de nucleosomas posicionados aguas arriba del TSS de esos mismos genes (Kumar y Bhargava, 2013).

-Modificadores de histonas:

Acetilinasas: una de las acetilinasas más importantes implicadas en la transcripción de RNAPIII es Rtt109, necesaria para la modificación de la histona H3 (Rufiange *et al.*, 2007).

Desacetilinasas: no se conoce muy bien su papel, pero se han encontrado las desacetilinasas Hda1, Hos2, Sir3 y Sir4 asociadas a los tRNAs *in vivo*.

4.3 Organización espacial de la transcripción

En *S. cerevisiae*, el gen del rRNA 5S (transcrito por RNAPIII) se localiza dentro del operón rDNA (locus RDN1), adyacente a los otros transcritos del rRNA (transcritos por RNAPI). El locus RDN1 es una región de aproximadamente 1-2 Mb que consiste en 100-200 copias de una repetición en tándem de 9.1Kb, localizada en el cromosoma XII. Esta agrupación lineal de genes ribosomales forma el nucléolo, una estructura tridimensional donde los rRNAs son transcritos, procesados y ensamblados en pre-ribosomas. Dado que el gen 5S forma parte por definición del nucléolo, la transcripción de la RNAPIII co-localiza en el nucléolo. Por el contrario, los genes de tRNA están repartidos a lo largo de los 16 cromosomas en levadura, sin formar agrupaciones u operones. A pesar de ello, se ha encontrado que un grupo grande de tRNAs se transcribe en el nucléolo, y que la enzima RNasa P, responsable de su maduración, también es de localización nucleolar (Thompson *et al.*, 2003). Se ha hipotetizado que la razón de esto es que la localización nucleolar permite que se produzca el efecto de silenciamiento dependiente de tRNAs (tgm, *tRNA gene mediated*), por el cual los genes de RNAPII adyacentes ven su expresión disminuida. Una excepción notable la constituyen los retrotransposones de la familia Ty (de los cuales se hablará más adelante), que se integran preferentemente muy próximos a tRNAs, pero que parecen haber evolucionado estrategias para adaptarse al silenciamiento.

4.4 Biogénesis de los tRNAs

Los tRNAs se transcriben como moléculas precursoras (pre-tRNA) que pasan por una elaborada serie de alteraciones post-transcripcionales para generar RNAs maduros (Hopper, 2013). Estas alteraciones incluyen la eliminación de nucleótidos en los extremos 5' y 3', la adición de nucleótidos a todos los extremos 3' y al extremo 5' de un tRNA, la eliminación de intrones, y la modificación de bases. Esta modificación consta de 25 tipos distintos de metilaciones, desaminaciones, isomerizaciones, y otras modificaciones. A continuación se enumeran brevemente los factores involucrados en estos pasos de procesado:

-Eliminación de secuencias en 5' y 3': La eliminación de los aproximadamente 12 nucleótidos extra que se encuentran en el extremo 5' del pre-tRNA la lleva a cabo la RNasa P, mientras que el extremo 3' es procesado en dos pasos. El primero requiere la exonucleasa 3'-5' Rex1, y el segundo la endonucleasa Trz1 (Ozanick *et al.*, 2009; Daoud *et al.*, 2012).

-Adiciones de bases en los extremos 3' y 5': todos los tRNAs contienen un triplete CCA en el extremo 3' necesario para su aminoacilación. La responsable de esta adición es la nucleotidil transferasa de tRNAs CCA1. Además de en 3', hay un tRNA que necesita un paso adicional de adición de una G para la adecuada maduración de su extremo 5' antes de ser aminoacidilado, el tRNA^{His}. Esta catálisis la realiza Thg1 (Gu *et al.*, 2003).

-Splicing de intrones: a pesar de que no se conoce muy bien la razón de su presencia, hay 60 genes de tRNAs que poseen intrones en *S. cerevisiae*. Estos intrones han de ser eliminados antes de que las moléculas puedan funcionar en la síntesis de proteínas. Los intrones están presentes en 10 familias de tRNAs distintas, se localizan siempre una base aguas abajo del triplete anticodón, y tienen una longitud de entre 14 y 60 nt. La eliminación de intrones se da en la periferia de la membrana mitocondrial en tres etapas: 1) eliminación del intrón en sí, catalizada por el heterotetrámero Sen2, Sen34, Sen15 y Sen54, 2) ligación de

los exones 5' y 3', catalizada por la ligasa Trl1 y, 3) eliminación del fosfato residual 2', catalizada por la fosfotransferasa Tpt1.

-Modificaciones de bases: más del 15% de los nts están modificados en los tRNAs de levadura. La distribución de las modificaciones ha sido compilada en <http://modomics.genesilico.pl/sequences/list/tRNA>. Las modificaciones son necesarias para diversas funciones como: discriminación de tRNAs, fidelidad traduccional, mantenimiento de la pauta de lectura y estabilidad de los tRNAs. A pesar de ello, muchos de los genes que codifican las proteínas responsables de las modificaciones no son esenciales (Phizicky y Hopper, 2010).

4.5 Sesgo en el uso de codones y selección traduccional.

El sesgo en el uso de codones hace referencia a la frecuencia de aparición de codones sinónimos en el DNA. Los codones o tripletes codifican para un aminoácido específico de la cadena polipeptídica o para la terminación de la traducción (codones de parada). Hay 64 codones diferentes (61 codifican aminoácidos y 3 codones de parada), pero sólo 20 aminoácidos. Esta sobreabundancia en el número de codones permite que muchos aminoácidos sean codificados por más de un codón. Es por esta redundancia por la que se dice que el código genético es degenerado. Diferentes organismos muestran preferencias particulares por uno de los diferentes codones que codifican el mismo aminoácido. Esto llevó a postular la hipótesis de que cada genoma tiene una *firma* de uso de codones particular que refleja las fuerzas evolutivas que están actuando sobre él (Grantham *et al.*, 1980). Está mayoritariamente aceptado que las preferencias por ciertos codones reflejan un balance entre sesgos mutacionales y selección natural para optimizar el proceso de la traducción. Los codones más usados (llamados óptimos, favoritos o preferidos) en organismos de crecimiento rápido, como *E. coli* o *S. cerevisiae*, reflejan la composición de su repertorio de tRNAs. Esto se conoce como el *paradigma levadura-E. coli*, siendo la fuerza evolutiva responsable de esto la llamada *selección traduccional* (dos Reis *et al.*, 2004). Se cree que la utilización de codones óptimos ayuda a alcanzar tasas de traducción más altas (Man y Pilpel, 2007) y mayor fidelidad (Drummond y Wilke, 2008). Como

resultado de estos factores, se espera mayor selección traduccional en los genes más expresados, como es el caso de los organismos mencionados. Por el contrario, en humanos y otros organismos, la selección traduccional está ausente, y las preferencias por unos u otros codones están determinadas por los sesgos mutacionales del genoma (contenido en G+C).

La naturaleza del sesgo en el uso de codones es un área de debate abierta. No está claro si el uso de codones es el que determina la evolución del repertorio de tRNAs, o viceversa. Se han desarrollado modelos matemáticos por los cuales tanto el uso de codones por los genes y la expresión de los tRNAs co-evolucionan retroalimentadamente. Esto es: los codones que ya están presentes en el genoma en una frecuencia alta dictan la expresión de sus tRNAs correspondientes, y los tRNAs que normalmente se expresan mucho dictan la frecuencia de aparición en los genes de los codones que reconocen. También se ha descubierto que existe un perfil conservado universalmente de eficiencia de traducción a lo largo de los mRNAs, por el que los primeros 30-50 codones serían traducidos por tRNAs sub-óptimos, y por tanto de manera poco eficiente o lenta, y los 50 codones últimos serían los más eficazmente traducidos (Tuller *et al.*, 2010). La razón de esta fase lenta inicial sería un último punto de regulación de las fases inicio de traducción/elongación temprana (Tuller *et al.*, 2010).

4.6 Elementos transponibles asociados a tRNAs.

El genoma de *S. cerevisiae* alberga cinco familias de retrotransposones tipo LTR (Ty1 a Ty5), estructural y funcionalmente relacionados con los retrovirus. Se han documentado más de 300 inserciones de estos elementos, lo que en conjunto supone un 3% de las 12 MB del genoma (Kim *et al.*, 1998). La transcripción de estos elementos es dependiente de RNAPII, y en su proceso de maduración se les añade caperuza en el extremo 5' y cola Poli(A) en 3'. A pesar de que la mayoría de ellos posee una TR baja (como resultado de promotores débiles e inserción en zonas heterocromáticas del genoma (Morillon *et al.*, 2002)), los mRNAs de los transposones de la familia Ty1 son

de los más abundantes de la célula, llegando a representar un 0.8% del RNA total de la levadura, lo cual se explica por su estabilidad, excepcionalmente alta (Curcio y Garfinkel, 1992).

Se ha descrito también que los transposones Ty tienen preferencia por insertarse en la vecindad de genes de la RNAPIII, sean tRNAs o no (Bolton y Boeke, 2003, Baller *et al.*, 2012). A pesar de ello, son pocos los trabajos que han estudiado una posible interacción entre las maquinarias de la RNAPII y la RNAPIII que resultara en una expresión diferencial de los genes de tRNAs. A nivel de unos pocos genes individuales, se ha descrito que no existe un gran efecto de la expresión de los transposones en la expresión de los tRNAs adyacentes (Kinsey y Sandmeyer, 1991, Bolton y Boeke, 2003).

Por otra parte, se ha descrito que los transposones tienen preferencia por insertarse en regiones ricas en nucleosomas posicionados que flanquean genes, y que por tanto se benefician de la estructura nucleosomal particular de las regiones tDNA, que obedece a tal estructura (ver figura 36, B, y Baller *et al.*, 2012). Esta estrecha relación entre los tRNAs y los transposones ha servido para hipotetizar que los genes de tRNA podrían dispersarse por el genoma gracias a la retrotranposición (Weiner *et al.*, 1986).

OBJETIVOS

Esta tesis parte de la existencia de una técnica genómica para el estudio de la transcripción naciente en levadura ampliamente utilizada y contrastada: el *Genomic run-on*, basada en la utilización de macrochips de las ORFs completas del genoma de *S. cerevisiae*. Debido a la aparición progresiva de nuevas plataformas que permiten interrogar la totalidad de las regiones del genoma, y a una resolución mayor, como los microchips de embaldosado o *tiling arrays*, el objetivo principal de esta tesis es la puesta a punto de una técnica adaptada a ellas que permita un análisis detallado de la transcripción naciente. Los objetivos concretos que se marcaron fueron:

- Desarrollar un nuevo procedimiento de run-on a escala genómica que sustituya el uso de las plataformas basadas en marcajes radiactivos y aproveche las plataformas de más resolución, así como de las herramientas bioinformáticas necesarias para el análisis de los datos generados.
- Estudiar los perfiles globales de transcripción naciente aprovechando el carácter específico de hebra de los datos para estudiar las dinámicas del transcriptoma global de levadura. Comparar y evaluar la complementariedad de los datos con otras medidas alternativas de tasas de transcripción existentes en la actualidad.
- Aplicar la técnica al estudio del efecto de mutantes relacionados con el ciclo de síntesis y degradación del RNA para poder extraer información sobre el funcionamiento de la maquinaria transcripcional y su regulación.
- Detectar posibles patrones de actividad de las RNA Polimerasas a lo largo de los transcritos y de las zonas flanqueantes que pudieran obedecer a condicionantes impuestas, tanto por su contexto cromatínico, como por otros factores.
- Caracterizar la transcripción naciente producida por las otras RNAP nucleares de levadura.
- Desarrollar un protocolo que permita analizar los RNAs nacientes a la máxima resolución mediante secuenciación masiva.

MATERIALES Y MÉTODOS

1. Materiales de especial relevancia

Enzimas, proteínas y reactivos

H6-GST-MBD (Diagenode, #mbd-001-050). Proteína de fusión para el aislamiento selectivo de RNAs marcados con uridinas metiladas. Consiste en la fusión en C-Terminal de la Glutación-S-Transferasa (GST) con el dominio de unión a grupos metilo (*Methyl Binding Domain*, MBD) de la proteína humana MeCP2, portando una etiqueta His6 (H6).

Streptavidin-HRP (Pierce, #N100). Estreptavidina acoplada a la enzima peroxidasa (*horseradish peroxydase*, HRP). Sistema utilizado para la purificación y/o detección selectiva de RNAs con uridinas biotiniladas

EZ-Link Biotin-HPDP (Thermo, #21341). Compuesto que reacciona con moléculas que contienen grupos tiol (-SH) formando puentes disulfuro reversibles. Utilizado para acoplar biotina reversiblemente a RNAs que han incorporado 4-Tio-U.

Anticuerpos

BrdU, (IIB5) (Santa Cruz Biotechnology, Inc.). Anticuerpo monoclonal que reconoce los residuos de bromo incorporados en el RNA naciente en forma de Bromo-UTP. Se ha utilizado libre en solución (sc-32323), como en su forma conjugada con partículas de agarosa para inmunoprecipitación (IP) (sc-32323 AC)

Anti-Bromodeoxyuridine (Roche, #11 170 376 001). Anticuerpo monoclonal que reconoce los residuos de bromo incorporados en el RNA naciente en forma de Bromo-U.

5-Methyl-Cytidine antibody (Abcam, #ab51552, cedido por Luis Torres, Univ. De Valencia). Anticuerpo monoclonal que reconoce grupos metilos incorporados en el RNA naciente en forma de Metil-C.

Inhibidores

Actinomicina D (Sigma, #A1410-2MG): inhibidor de la replicación y de la transcripción. En este estudio se ha utilizado como método para evitar la síntesis no deseada de la segunda hebra de cDNA en reacciones de transcripción inversa (Perocchi 2007).

α -amanitina (Sigma, #A2263-1MG): inhibidor específico de la RNAPII

ML-60218 (Calbiochem, #557403-10MG): Inhibidor específico de RNAPIII.

Nucleótidos modificados

5-Methyl-CTP (TriLink Biotech, #N-1014-041301)

5-Bromo-UTP (Sigma, # B7166-5MG)

4-Thio-UTP (TriLink Biotech, # N-1025)

Biotin-11-UTP (Ambion, #AM8451)

Otros

Sarkosil. El sarkosil es un surfactante aniónico que permeabiliza las membranas de las células de levadura y desorganiza la estructura cromatínica del DNA (Hawley y Roeder 1985). El sarkosil bloquea nuevos inicios de transcripción pero preserva intactas las estructuras ternarias entre DNA, RNAPs, y RNAs nacientes.

Glutathione–Agarose (Sigma, #G4510). Perlas de agarosa acopladas a glutatión, utilizadas para unir específicamente proteínas con dominios GST.

Glucógeno para biología molecular (Roche, #10 901 393 001). Polisacárido inerte. Dado que es insoluble en etanol e isopropanol, añadido a las muestras de RNA en el paso de precipitación forma un precipitado que arrastra a los ácidos nucleicos, aumentando el rendimiento de las precipitaciones. Además forma un agregado blanquecino al fondo del tubo, lo que ayuda a la visualización de *pellets* de RNA pequeños.

2. Técnicas microbiológicas

2.1 Cepas de levadura

Tabla 1. Cepas de *S. cerevisiae* usadas en este trabajo.

Cepa	Genotipo Relevante	Procedencia
BQS252	Mat a, <i>ura3-52</i> , procede de FY1679 (fondo genético S288c)	Susana Rodríguez Navarro, tesis doctoral.
BQS408	Mata, <i>rpb4Δ::KanMX4</i>	Sebastián Chávez
BQS409	<i>dst1Δ::KanMX4</i>	Sebastián Chávez
YOR304W	<i>isw2Δ::KanMX4</i> (fondo genético BY4741)	Vicent Pelechano (<i>Stanford Gene Deletion Project</i>)
yMC458	Mat a, <i>xrn1Δ</i> , <i>ade2</i> , <i>ura3-52</i>	Mordechai Choder

2.2 Medios y condiciones de cultivo

Para crecer las levaduras se usó el medio rico YPD. Este medio contiene extracto de levadura 1%, peptona bacteriológica 2% y glucosa 2%. Los cultivos se crecieron en matraces con una agitación de 190 rpm y a 28°C.

Las bacterias se crecieron en medio rico LB (extracto de levadura 0.5%, bacto-triptona 1%, NaCl 1%), o medio LBA (medio LB suplementado tras el autoclavado con el antibiótico Ampicilina, 100 µg/L). Los cultivos se crecieron con una agitación de 190 rpm y a 37°C.

Los medios sólidos se prepararon añadiendo un 2% de agar antes del autoclavado.

Todos los medios se esterilizaron autoclavándolos durante 20 minutos a 121°C.

3. Técnicas de Biología Molecular

3.1 Obtención de ácidos nucleicos

3.1.1 Consideraciones previas.

Preparación de RNasa A libre de DNasas. Para garantizar la integridad del DNA en las preparaciones, en esta tesis se ha trabajado con RNasa A libre de DNasas. Se procedió tal y como se describe en el manual de laboratorio *Molecular Cloning*, apéndice A4.39 (Sambrook y Russell, 2001). La preparación consistió en disolver la RNasa A (Roche) a una concentración de 10 mg/mL en 0.01 M de acetato sódico pH 5.2 y calentarla a 100°C durante 15 min. Tras ello se dejó enfriar lentamente hasta alcanzar temperatura ambiente y se ajustó el pH añadiendo 0.1 volúmenes de Tris-HCl pH 7.4. Se alicuotó y se conservó a -20°C.

Agua para RNA. En todos los experimentos en los que se trabajó con RNA se utilizó agua destilada libre de RNasas. Para ello se autoclavó durante 1 hora a 134°C y se guardó en alícuotas de 1 mL en tubos *Eppendorf*.

3.1.2 Extracción de DNA de levadura

Por cada experimento de extracción de DNA se emplearon 100 mL de un cultivo de levadura crecido hasta la mitad de la fase exponencial $DO_{600} = 0.5$. Las células se recogieron por centrifugación y, tras eliminar el medio de cultivo, se resuspendieron en 500 μ L de tampón *10prep* (2% TritonX-100, 1% SDS, 100 mM NaCl, 1 mM EDTA y 10 mM Tris-HCl pH 8) y se transfirieron a un tubo de rosca cónico conteniendo 500 μ L de una mezcla de fenol:cloroformo:alcohol isoamílico (25:24:1). La rotura de las células se consiguió con un agitador automático FastPrep (MP), programado con tres pulsos de 30 seg a una intensidad de 5.5 m/s. Tras la rotura se separó la fase acuosa de la orgánica

mediante centrifugación. La fase acuosa se mezcló de nuevo con 500 μ L de fenol:cloroformo:alcohol isoamílico (25:24:1), se agitó durante 1 min, y se volvió a recuperar tras centrifugación. Para eliminar los posibles restos de fenol se limpió la muestra con un pase de 500 μ L de una mezcla de cloroformo:alcohol isoamílico (24:1), y separación de la fase acuosa por centrifugación. El DNA se precipitó añadiendo 1/10 de NaAc 3 M y 2 volúmenes de EtOH 96% durante toda la noche a -20°C. Al día siguiente el DNA precipitado se lavó con EtOH 70% y se resuspendió en 400 μ L de tampón TE (1 mM EDTA, 10 mM Tris HCl pH 8). Para eliminar posibles restos de RNA coextraídos se incubó la muestra durante 30 min a 37°C con 3 μ L de RNasa A (10 mg/mL, libre de DNasa).

El DNA obtenido se cuantificó espectrofotométricamente con NanoDrop (Thermo) y se corrió en un gel de agarosa para comprobar su integridad.

3.1.3 Extracción de RNA de levadura

Para el aislamiento del RNA de las cepas de levadura se utilizaron dos metodologías distintas, dependiendo de si los RNAs a aislar estuvieran o no marcados con Biotina. El aislamiento de RNAs sin biotinar se realizó mediante un protocolo de extracción que incluye un paso de fenolización, detallado a continuación:

3.1.3.1 Método del fenol ácido. Para extraer RNA se partió de 100 mL de un cultivo de levadura crecido hasta la mitad de la fase exponencial $DO_{600} = 0.5$. Las células se recogieron por centrifugación, se resuspendieron en 500 μ L de tampón LETS (0.1M LiCl, 10 mM EDTA, 0.2% SDS y 10 mM Tris-HCl pH 7.4) y se transfirieron a un tubo de rosca cónico conteniendo 500 μ L de una mezcla de fenol:cloroformo:alcohol isoamílico (25:24:1). La rotura de las células se consiguió con un agitador automático FastPrep (MP), programado con tres pulsos de 30 seg a una velocidad de 5.5 m/s. Tras la rotura se separó la fase acuosa de la orgánica mediante centrifugación. Para eliminar los posibles restos de fenol se limpió la fase orgánica con un pase de 500 μ L de una mezcla de cloroformo:alcohol isoamílico (24:1), y se recuperó la fase acuosa tras centrifugación. El

RNA se precipitó añadiendo 1 volumen de LiCl 5M durante toda la noche a -20°C. Al día siguiente el RNA precipitado se lavó con EtOH 70% y se resuspendió en 400 µL de agua. Seguidamente, y con el objetivo de eliminar los restos de litio de la primera precipitación, se volvió a precipitar con 1/10 de NaAc 3 M y 2.5 volúmenes de EtOH 96%. El RNA se limpió de DNA añadiendo DNasa (libre de RNasa, Roche), se volvió a fenolizar, y se resuspendió en 300 µL de tampón TE. El RNA obtenido se cuantificó espectrofotométricamente y se corrió en un gel de agarosa para comprobar su integridad.

3.1.3.2 Método de la Proteinasa K. Dado que es sabido que las moléculas biotiniladas son arrastradas junto con el fenol a la fase orgánica (Vicent Pelechano, Tesis Doctoral), no es posible hacer una purificación de los RNAs marcados con biotina generados mediante BioGRO (detallado en sección *Genomic run-on* no radiactivo), mediante la técnica descrita en el apartado anterior. En esos casos se realizó una extracción de RNA independiente de fenol basada en el kit de extracción de RNA total *MasterPure Yeast RNA Purification Kit* (Epicentre Biotechnologies), específicamente:

A partir de un cultivo igual al descrito en el método anterior, se recogieron las células y se resuspendieron en 300 µL de tampón de extracción de RNA, al que se le añadió Proteinasa K (Roche) para que quedara a una concentración final de 0.16 µg/µL. La mezcla se incubó durante 15 min a 70°C, agitando en *vortex* la muestra cada 5 min para evitar que las células se agregaran en el fondo del tubo. Tras la incubación, la mezcla se enfrió en hielo 3-5 min, se le añadieron 175 µL del reactivo de precipitación de proteínas MPC y se mezcló vigorosamente durante 10 seg. La mezcla se centrifugó durante 10 min a 4°C en una microcentrífuga de mesa y se transfirió el sobrenadante a un nuevo tubo. El RNA se precipitó durante al menos 3 horas a -20°C añadiendo 500 µL de isopropanol. Adicionalmente se añadió 1 µL de glucógeno (5 µg/µL, Roche) para incrementar el rendimiento de la precipitación. Tras la precipitación, el RNA se lavó con etanol al 70% y se resuspendió en 87 µL de agua de RNA. A diferencia del método del fenol ácido, el método de la Proteinasa K no discrimina entre DNA y RNA, con lo que la resuspensión final contiene una mezcla de ambos ácidos nucleicos, en proporciones desconocidas. Por tanto, para eliminar los restos de DNA que pudieran interferir en etapas posteriores del procesado se llevó a cabo una digestión enzimática. Para ello se trató la muestra con 2 µL

de DNasa I, libre de RNasas (Roche) en 10 μ L de tampón de reacción 10X (400 mM Tris-HCl, 100 mM NaCl, 60 mM MgCl₂, 10 mM CaCl₂; pH 7.9) y se añadió 1 μ L de inhibidor de RNasas (*RNase OUT*, Invitrogen) para preservar la integridad del RNA. La mezcla se incubó a 37°C durante 30 min, tras los cuales se elevó el volumen de la mezcla hasta los 200 μ L con agua y se purificó siguiendo las instrucciones del kit. El RNA purificado se resuspendió en 32 μ L de agua y se procedió a su cuantificación espectrofotométrica.

3.1.4 Extracción de DNA plasmídico de *E. coli*

En esta tesis se ha trabajado con dos plásmidos replicados en *E. coli*, los cuales portan insertos con la secuencia de fragmentos de dos genes distintos de la bacteria *Bacillus subtilis*. Los plásmidos proceden de la Colección Americana de Cultivos Tipo (ATCC) y sus referencias son: pGIBS-LYS (ATCC Number 87482) y pGIBS-PHE (ATCC Number 87483). A partir de un cultivo de *E. coli* crecido en medio LBA se procesaron 1.5 mL para extraer el DNA plasmídico mediante el kit *Perfectprep Plamid Mini* (Eppendorf), siguiendo las instrucciones del fabricante.

3.2 Técnicas electroforéticas de separación de ácidos nucleicos

3.2.1 Electroforesis nativa en gel de agarosa

Los ácidos nucleicos se separaron por tamaño en geles horizontales de agarosa al 1% para DNA y al 1.8% para RNA en tampón TAE (40 mM Tris-acetato, 1 mM EDTA). Se prepararon añadiendo BrEt 5 μ g/mL y se iluminaron con luz ultravioleta para revelar la posición de los ácidos nucleicos una vez transcurrida la electroforesis.

3.2.2 Electroforesis en gel desnaturizante de agarosa

Los RNAs se separaron por tamaños usando geles desnaturizantes con 1.2% de agarosa en tampón MOPS 1X (20 mM MOPS, 4 mM NaAc, 1 mM EDTA, pH 7) con 0.064% de formaldehído. La preparación de las muestras consistió en mezclar 10 µg de RNA con una disolución de formaldehído al 6.5%, formamida al 50%, MOPS 1X, tampón de carga de RNA (30% glicerol y 0,25% azul de bromofenol) y BrEt 5 µg/mL. Posteriormente, se desnaturizaron las muestras calentándolas durante 10 minutos a 65°C y enfriándolas en hielo. Una vez cargadas las muestras, se dejó correr durante 3 horas a 70 V. Finalmente los RNAs se visualizaron mediante luz ultravioleta.

3.2.3 Electroforesis en gel desnaturizante de poliacrilamida

Los RNAs se separaron por tamaño en minigeles desnaturizantes de poliacrilamida al 8%, con una concentración de urea de 7 M en TBE 1X (1 mM EDTA, 45 mM Tris-Borato pH 8). Para la preparación de las muestras se calentó 1 µg de RNA en un volumen de 10 µL a 70°C durante 5 min y se pasó a hielo. Lo mismo se hizo para la preparación del marcador de tamaño *Low Range ssRNA Ladder* (New England Biolabs, NEB), del que se cargaron 10 µL (500 ng/µL). Se utilizó el sistema de electroforesis de minigeles MiniProtean (BioRad), que se programó para que corriera a 160 V durante 45 min. Seguidamente se tiñó el gel durante 15 min sumergiéndolo en 2 mL de una disolución de BrEt en agua destilada (1:1000). Finalmente los RNAs se visualizaron mediante luz ultravioleta.

3.2.4 Electroforesis capilar en dispositivos compactos

Se utilizó el sistema de electroforesis capilar Experion (BioRad) para separar por tamaño y cuantificar muestras de RNA. Para ello se siguieron las instrucciones del fabricante detalladas en el kit para la cuantificación de RNAs *Experion RNA StdSens Analysis Kit*. Tras

poner a punto el aparato y el chip de electroforesis, se prepararon las muestras para su carga: primero se eliminaron las estructuras secundarias de 0.5 µg de cada RNA de interés y de 2 µL del marcador de tamaño del kit incubando a 70°C durante 5 min y pasando a hielo. Seguidamente se procedió a cargar tanto el marcador de tamaño como las muestras de RNA en el chip y se ajustaron los parámetros de electroforesis mediante el software del Experion, seleccionando el programa *Eukaryotic Total RNA Assay with RNA Quality Indicator (RQI)*.

3.3 Reacción en cadena de la polimerasa (PCR)

3.3.1 PCR convencional para generar moldes para IVT

De manera genérica se realizaron amplificaciones de regiones concretas del DNA mediante la reacción de PCR bajo las siguientes condiciones: en un volumen final de 100 µL se mezclaron 2.5 U de DNA Polimerasa termoestable (1 U/µL, Biotools), tampón de reacción 10X (Tris HCl 750 mM pH 9.0, KCl 500 mM, (NH₄)₂SO₄ 200 mM), 3mM de MgCl₂, 0.2 mM de dNTPs, 0.5 µM de cada cebador y 500 ng de DNA molde.

La PCR se llevó a cabo indistintamente en los termocicladores GeneAmp PCR System 9700 (ABI) y Mastercycler Personal (Eppendorf). El perfil térmico consistió en un primer paso de desnaturalización a 94°C durante 3 min, seguido de 25 ciclos de una sucesión de etapas de desnaturalización a 94°C durante 30 seg, anillamiento de los cebadores a 52°C durante 30 seg, y extensión a 72°C durante 1 minuto. El último paso consistió en una extensión final de las cadenas sintetizadas a 72°C durante 5 min.

Para cubrir los objetivos concretos de otros experimentos se modificó el diseño general de la reacción de PCR tanto a nivel de los componentes de la mezcla de reacción, como a nivel del perfil térmico. Cada una de estas modificaciones se detalla a lo largo de la tesis en el apartado correspondiente.

3.3.2 PCR cuantitativa (qPCR) de comprobación de enriquecimientos selectivos

Para comprobar la eficiencia de los métodos de purificación selectiva de RNAs marcados, se utilizó la técnica de la qPCR para detectar y cuantificar la presencia final de los RNAs sintéticos inyectados como acompañantes de las muestras de RNAs para purificación selectiva. Todos los experimentos de qPCR se realizaron con el termociclador DNA Engine acoplado al detector Chromo4 (BioRad) y se analizaron usando el programa *Opticon Monitor* (BioRad). Para las amplificaciones se usó la mezcla de qPCR *Quantimix Easy SYG* con Rox (Biotools) basada en la detección del producto amplificado por el fluoróforo *SYBR Green*.

El perfil térmico consistió en una desnaturalización inicial de 1 min a 95°C seguidos por un total de 45 ciclos consistentes en: 30 seg desnaturalización a 95°C, 30 seg de apareamiento de oligos a 55°C y 30 seg de elongación a 72°C. En cada ciclo se midió la fluorescencia después de la elongación a 75°C, 80°C y 82°C, dependiendo de la temperatura de fusión del producto amplificado. Finalmente se realizó una curva de desnaturalización de los productos de PCR. Para cada reacción de PCR se utilizaron los datos de fluorescencia obtenidos a la mayor temperatura que no desnaturalizara el producto de PCR.

3.3.3 Oligonucleótidos

Tabla 2. Oligonucleótidos utilizados en este trabajo.

Nombre	Secuencia
pGibs-LYS-D	5'-TCT TGG AGG AGG TTT CGG CAT TC-3'
pGibs-LYS-R	5'-CCG TAT AAA GAG TTG TGC CTG CGT C-3'
pGibs-PHE-D	5'-TGC CGC AGG ATG ATC AGT CC-3'
pGibs-PHE-F	5'-TGC AGC CGA GTG CTT CAA GC-3'
5ETS-D	5'-GAG AGT CGC TAG GTG ATC-3'
5ETS-R	5'-GTA CGT TTG ATA TCG CTG-3'
RPS3-3D	5'-GCT GAC GGT GTC TTC TAC GC-3'
RPS3-3R	5'-TCT TCT GAT AGC CAA ACC GTT C-3'
T3	5'-ATT AAC CCT CAC TAA AGG GA-3'

T7	5'-TAA TAC GAC TCA CTA TAG GG-3'
Oligo d(G) HN	5'-GGG GGG GGG GGG GGG HN-3'
GX1-ILLUMINA	5'-CAA GCA GAA GAC GGC ATA CGA-3'
GX2-T7	5'-TAA TAC GAC TCA CTA TAG GGG TTC AGA GTT CTA CAG TCC GA-3'
Tiling Adapters Illumina	5'-GGG TTC AGA GTT CTA CAG TCC GAC GAT CNN NNN NNN NNN NTC GTA TGC CGT CTT CTG CTT-3'
Multiplex 3' SR Adaptor	5'-rAppAGATCGGAAGAGCACACGTCT-NH2-3'
Multiplex 5' SR Adaptor	5'- rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-3'
Multiplex SR RT Primer	5'-AGA CGT GTG CTC TTC CGA TCT-3'

(tabla 2, continuación)

3.4 Transferencia de ácidos nucleicos a membranas

3.4.1 Dot Blot

La técnica de *dot blot* se empleó para las comprobaciones del rendimiento de algunos de los procesos de enriquecimiento de RNAs marcados con biotina. Como control positivo de marca se utilizó un RNA sintetizado en presencia de únicamente Bio-11-UTP, y como control negativo un RNA sintetizado en presencia de sólo UTP. Tanto los controles como las muestras se prepararon a una concentración de 30 ng/μL en agua y se realizaron ocho diluciones seriadas. De cada dilución se aplicó 1 μL sobre una membrana de nailon cargada positivamente *Hybond N+* (GE Healthcare). Una vez depositados los RNAs, la membrana se dejó secar al aire y se fijó mediante irradiación con 50 mJ de luz ultravioleta con un entrecruzador GS Gene Linker (BioRad). A partir de este punto, la membrana se mantuvo todo lo posible en oscuridad para minimizar la posible fotodegradación de la biotina. La membrana se bloqueó con tampón de bloqueo (10% SDS, 125 mM NaCl, 7 mM Fosfato Monosódico, y 9 mM fosfato disódico) durante 20 min a temperatura ambiente en agitación, se incubó con estreptavidina-hrp (1:4000) durante 10 min, se lavó dos veces con tampón de bloqueo, y dos veces más con tampón de lavado (0.1 M Tris-HCl pH 9.5,

0.1 M NaCl, 10 mM MgCl₂). Para el revelado se utilizó el sistema quimioluminiscente ECL (GE Healthcare), y la señal se capturó mediante películas de autoradiografía *XFilm* (Fuji).

3.4.2 Northern Blot a partir de geles de agarosa

Tras la electroforesis en gel de agarosa, tanto en condiciones nativas como desnaturizantes, el RNA se transfirió a membranas de nailon cargadas positivamente mediante la técnica de *Northern Blot*, tal y como se describe en el manual de laboratorio *Molecular Cloning*, Capítulo 7 (Sambrook y Russell 2001), basada en el paso de los ácidos nucleicos desde los geles a las membranas mediante capilaridad en presencia de tampón SSC 6X (0.9 M NaCl y 90 mM citrato sódico) durante toda la noche. El RNA se entrecruzó tratando las membranas con 50 mJ de radiación ultravioleta.

Debido a que en este estudio se obtuvieron RNAs nacientes marcados mediante la técnica del run-on con nucleótidos modificados de distinta naturaleza (UTP-radiactivo, UTP-biotina), el procesamiento y el revelado de la señal fue distinto en cada caso. El revelado de la señal radiactiva consistió en el sellado de la membrana con plástico para uso alimentario y su exposición en una pantalla IP *BAS-MP* (FujiFilm). La adquisición de imágenes se llevo a cabo con un lector FLA3000 Phosphorimager (FujiFilm). El procesamiento de la membrana para la detección de los RNAs biotinilados siguió el mismo proceso que las membranas utilizadas en el apartado anterior (*Dot Blot*), mientras que la imagen se obtuvo mediante el sistema ImageQuant LAS 4000 mini (GE).

3.4.3 Northern Blot a partir de geles de poliacrilamida

Al igual que en apartado anterior, los geles de poliacrilamida sirvieron para separar por tamaño tanto RNAs marcados radiactivamente como RNAs biotinilados, por lo que el procesamiento de los geles se ajustó de acuerdo a ello.

Los geles de muestras radiactivas se fijaron sumergiéndolos en una solución de ácido acético glacial al 7% y se dejaron en agitación suave durante 30 min. Tras ello se colocaron sobre papel Whatmann, se cubrieron con plástico para uso alimentario y se desecaron en un desecador de geles SGD5040 Slab Gel Dryer (ThermoSavant) durante 1 hora a temperatura ambiente. Una vez secos se expusieron en una pantalla IP y se adquirió la imagen.

Los geles de muestras biotiniladas se transfirieron a membranas de nailon mediante el método *Wet* con el sistema MiniProtean Tetra System (BioRad) en TBE 0.5X durante 90 min a 100 V en la cámara fría. Una vez transcurrida la transferencia las membranas se secaron entre papeles, se entrecruzaron con luz UV y se procesaron de la misma manera que en las transferencias por *Northern Blot*.

3.5 Biotinilación reversible de RNAs tiolados

Para acoplar moléculas de biotina a los RNAs marcados con Tio-U mediante *run-on* se siguió el protocolo *RNA Labelling* de Lars Doelken (http://www.bio.ifi.lmu.de/webfm_send/2309), con modificaciones: se marcaron 50 µg de RNA de *run-on*, y para ello se emplearon 2 µL de Biotin-HPDP (1 mg/mL en Dimetilformamida) por cada microgramo de RNA, 1 µl de tampón de biotinilación (10 mM Tris pH 7.4, 1 mM EDTA) por microgramo de RNA y 7 µL de H₂O de RNA por µg de RNA. La reacción transcurrió en tubos *Eppendorf* durante 90 min a temperatura ambiente en oscuridad en un agitador de rotación. Después se dividió la muestra en dos tubos; en el primero se precipitó el RNA con un 1/10 de NaCl 5M más un volumen de isopropanol a -20°C durante toda la noche, y en el segundo se eliminó el compuesto Biotin-HPDP no incorporado mediante extracción con un volumen de cloroformo-alcohol isoamílico (24:1). El RNA se precipitó con 1/10 de NaCl 5M más un volumen de isopropanol a -20°C durante toda la noche. Al día siguiente ambos RNAs se recogieron por centrifugación, se lavaron con alcohol al 70%, se resuspendieron en 20 µL de agua, y se cuantificaron espectrofotométricamente.

Los RNAs biotinilados por este procedimiento pueden ser liberados de la molécula de biotina mediante incubación con agentes reductores como el DTT 0.1 M durante 10 minutos en agitación a temperatura ambiente.

3.6 Aislamiento específico de RNAs marcados

3.6.1 IP de RNA con Bromo-U

En este trabajo se han utilizado dos estrategias para purificar de manera selectiva RNAs obtenidos por run-on en presencia de Br-UTP, que difieren tanto en el anticuerpo utilizado, como en la columna de purificación.

IP con anticuerpos unidos a esferas magnéticas. Se procesaron 100 µg de RNAs obtenidos mediante run-on en presencia de Br-UTP por experimento. Con el objetivo de poder monitorizar por PCR cuantitativa al final del proceso el grado de selectividad de los anticuerpos frente a RNAs con bromo se añadieron a las muestras cantidades conocidas de RNAs sintéticos generados en presencia, o bien del nucleótido modificado, o del nucleótido natural, en proporciones de 1000:1. De esa mezcla se separó una alícuota para la posterior qPCR.

Se prepararon dos juegos de esferas magnéticas: conjugadas con anticuerpo, y sin conjugar. Para las esferas conjugadas, 0.6 µg del anticuerpo anti-BrdU (0.1 µg/µL, Roche) se mezclaron con 50 µL de esferas magnéticas *Dynabeads Anti-Mouse IgG* (Invitrogen) lavadas. Se añadieron 100 µL de tampón PBS/BSA (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 5 mg/mL BSA, pH 7.4) y se dejó incubando en rotación 1 h a temperatura ambiente. A los 30 min de la incubación se añadieron 7 µg de DNA de salmón sonificado y desnaturalizado como agente bloqueante, y se dejó otros 30 min rotando a temperatura ambiente. Las esferas se lavaron con ayuda de un imán tres veces con 500 µL de PBS/BSA y se resuspendieron en 100 µL de PBS/BSA. La preparación de las

Dynabeads sin conjugar se hizo de la misma manera, con la única diferencia de que se añadió PBS/BSA en lugar del anticuerpo anti-BrdU a las esferas.

Cada muestra de RNA se dividió en dos tubos, uno que se incubó con las esferas conjugadas (muestra IP), y otro que se incubó con las esferas no conjugadas (control negativo, muestra *Mock*). Los tubos de RNA se desnaturalizaron a 80°C durante 10 min y se enfriaron en hielo 5 min. Seguidamente se les añadió 10 µL de inhibidor de RNasas y cada mezcla se incubó con un juego de *Dynabeads* distinto, llevándose en paralelo a partir de ahí. La incubación transcurrió durante 1 h en rotación a temperatura ambiente. Se recuperó el sobrenadante (no retenido) para qPCR, y las esferas se lavaron una vez con 500 µL de PBS/BSA y tres veces con 500 µL de PBS. Seguidamente se eluyeron dos veces los RNAs retenidos de cada muestra calentando a 80°C durante 10 min en 60 µL de PBS/SDS 1% en agitación vigorosa. De cada pase se recuperaron 50 µL de eluido, que se limpió de SDS utilizando columnas de exclusión molecular *ProbeQuant G-50* (GE).

Para la llevar a cabo la qPCR, tanto los no retenidos, como los eluidos purificados se trataron con DNasa I para eliminar restos del DNA bloqueante y se convirtieron a cDNA mediante RT (se detalla más adelante en el apartado *transcripción inversa*).

Inmunopurificación con anticuerpos unidos a esferas de agarosa. Debido a que las esferas de agarosa utilizadas para este experimento vienen preconjugadas con el anticuerpo BrdU (IIB5): sc-32323-AC (Santa Cruz Biotechnologies), para diseñar el experimento se prescindió de dividir la muestra de run-on en un IP y un *Mock*, procesándose como una única muestra a inmunoprecipitar.

Se siguió el esquema de IP descrito en (Core *et al.*, 2008), con modificaciones. El primer paso consistió en lavar 50 µL de bolas comerciales preconjugadas con anticuerpo con 800 µL de tampón de bloqueo (SSPE 0.5X, EDTA 1mM, Tween 20 0.05%, PVP 0.1%, BSA 1 mg/mL) y resuspenderlas en tampón de conjugación (SSPE 0.5X, EDTA 1mM, Tween 20 0.05%). La preparación de los RNAs de run-on fue tal y como se describe en el apartado anterior. Una alícuota se guardó para qPCR. La muestra se mezcló con las esferas y se mantuvo incubando en rotación a temperatura ambiente durante 1 h. Las esferas se recogieron por centrifugación a 4°C durante 30 seg a 3000 rpm. El sobrenadante se

recuperó y se guardó para qPCR. Las bolas se lavaron con un pase de 800 μ L de tampón *Low Salt Buffer* (SSPE 0.2X, EDTA 1mM, Tween 20 0.05%), dos pases de 800 μ L de tampón *High Salt Buffer* (SSPE 0.5X, EDTA 1mM, Tween 20 0.05%, NaCl 150 mM) y dos pases con tampón TET (TE 1X, Tween 20 0.05%). Los RNAs se eluyeron con cuatro pases de incubación con tampón de elución (DTT 20 mM, NaCl 300 mM, Tris-HCl 0.05 mM pH 7.5, EDTA 1 mM, SDS 0.1%) durante 5 min por pase a 65°C en agitación vigorosa.

Para la llevar a cabo la qPCR, tanto los No Retenidos, como los eluidos se convirtieron a cDNA mediante RT.

3.6.2 IP de RNA con Metil-C

La purificación selectiva de RNAs de run-on marcados con Metil-CTP se basó en el protocolo descrito en (Kangaspeska *et al.*, 2008). Se partió de 25 μ g de RNAs de run-on a los que se les añadieron transcritos sintéticos generados en presencia o bien de Metil-CTP o de UTP, en una proporción 1:1000. A la mezcla de RNAs de run-on y sintéticos se les añadió *Pull-Down Buffer*, PDB (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA) hasta alcanzar un volumen inicial de 530 μ L. De ellos se separaron 30 μ L para el análisis posterior por qPCR. A continuación se añadieron 0.1 μ L (1 μ g) de la proteína de fusión GST-MBD (Diagenode) y se mantuvo durante 2 horas a 4°C en rotación constante. Los complejos proteína-RNA se capturaron mediante la adición de 100 μ L de un *slurry* al 50% de perlas de agarosa recubiertas de glutatión (Glutathione-Agarose, Sigma) e incubación durante 1 hora más a 4°C. Tras recuperar la fracción No Retenida y reservar 30 μ L para posterior qPCR, se lavaron las esferas con cuatro pases de 500 μ L de PDB para eliminar las uniones inespecíficas. La elución se realizó añadiendo 500 μ L de glutatión 10 mM (Sigma) e incubando 10 min a temperatura ambiente. 30 μ L del eluido se procesaron para qPCR

Las tres alícuotas de 30 μ L para qPCR se purificaron mediante columnas de exclusión molecular *ProbeQuant G-50* antes de someterlas a RT y qPCR.

3.6.3 Purificación de RNA biotinilado con bolas magnéticas

Purificamos los RNAs biotinilados procedentes de run-on usando esferas magnéticas recubiertas de estreptavidina (*Dynabeads M-280 Streptavidin*). Las *Dynabeads* se prepararon cogiendo 50 μL de solución de bolas y lavándolas con tres pases de tampón *Binding and Washing 1X* (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl). Tras los lavados, las bolas se bloquearon con 100 μg de tRNA de *E. coli* (Sigma) durante 25 min en rotación a temperatura ambiente, tras lo cual se lavaron con tres pases de tampón *Binding and Washing 1X*. Una vez lavadas se resuspendieron en 55 μL de tampón *Binding and Washing 2X*. Las bolas lavadas y bloqueadas se mezclaron con 55 μL de RNAs biotinilados de run-on y la mezcla se incubó durante 30 min a temperatura ambiente en rotación y en condiciones de oscuridad. Se lavaron con tres pases de tampón *Binding and Washing 1X* y dos pases de tampón SSC 2X. Los RNAs se eluyeron de las bolas en 50 μL de agua calentando a 95°C durante 10 min con agitación vigorosa.

3.7 Transcripción inversa (RT)

Se copiaron los RNAs a cDNA mediante transcripción reversa partiendo de 11.3 μL de eluidos de las purificaciones selectivas, a los que se les añadieron 1.2 μL de cebadores aleatorios (3 $\mu\text{g}/\mu\text{L}$, Invitrogen) y 1 μL de inhibidor de RNAsas. Seguidamente se calentó la muestra a 72°C durante 10 minutos, dejándola a continuación enfriar a temperatura ambiente. Posteriormente se añadieron 6 μL de 5x *First Strand Buffer*, 3 μL de DTT 0.1 M, 1.5 μL de dNTPs (16 mM), y 1 μL de retrotranscriptasa *SuperScript II RT* (200 U/ μL , Invitrogen). Para realizar la RT se incubó la muestra durante 1 hora a 42°C más 15 min a 72°C para inactivar la enzima. Se eliminó el exceso de nucleótidos no incorporados aumentando el volumen de la muestra hasta 100 μL y usando una columna de purificación por exclusión molecular *S-300 HR* (GE).

3.8 Transcripción *in vitro* (IVT)

El experimento modelo de producción de RNAs sintéticos consistió en mezclar en un volumen final de 20 μL de reacción: 1 μg de DNA molde, 2 μL de tampón de reacción 10X comercial (Ambion), 20 U de Polimerasa del fago T7 (20 U/ μL , Ambion) y 10 mM de cada rNTP. Para los experimentos en los que se generaron RNAs con nucleótidos modificados, éstos se añadieron a una concentración de 1 mM, quedando los demás igual.

La reacción de IVT se llevó a cabo durante 16 h a 16°C en termociclador. Los transcritos producidos se limpiaron del DNA molde y los nucleótidos no incorporados con el kit de purificación *MasterPure Yeast RNA Purification Kit* (Epicentre Biotechnologies), siguiendo las instrucciones del fabricante.

Los productos finales purificados se cuantificaron espectrofotométricamente y su integridad se comprobó en un gel de agarosa al 1.8 %.

3.9 Selección de ácidos nucleicos por tamaño

Los RNAs de pequeño tamaño (<200 nt) se seleccionaron mediante el uso del kit de aislamiento de micro RNAs *NucleoSpin miRNA* (Macherey-Nagel), siguiendo las instrucciones del apartado 6.4 *Purification of siRNA and large dsRNA from DICER reactions*: a los RNAs disueltos en 150 μL de agua se les añadió un volumen de *Buffer ML* y se mezcló con *vortex* durante 5 seg. A la mezcla se le añadió 200 μL de EtOH 96%, se agitó con *vortex* 5 seg y se incubó a temperatura ambiente durante 5 min. Para separar los RNAs grandes de los RNAs pequeños la mezcla se cargó en una columna *NucleoSpin RNA Column* y se centrifugó a máxima potencia durante 30 seg. En la columna quedaron retenidos los RNAs grandes, mientras que en el líquido que pasó a través de ella quedaron los RNAs pequeños. La columna de los RNAs grandes se mantuvo en hielo hasta su posterior utilización. Para purificar los RNAs pequeños se añadieron 100 μL de *Buffer MP*, se mezcló con *vortex* 5 seg y se mantuvo 5 min a temperatura ambiente. Tras ello se

añadieron 800 μL de *Buffer MX* y se *vorteo* 5 seg, quedando la muestra en un volumen final de 1400 μL . 700 μL se cargaron en una columna *NucleoSpin miRNA Column* y se centrifugaron durante 30 seg a máxima potencia, descartándose el líquido que pasó a través de la columna. Tras ello se repitió el paso con el resto de la muestra. A partir de este punto tanto la columna con los RNAs grandes, como la columna con los RNAs pequeños se lavaron en paralelo añadiendo 700 μL de *Buffer MW2* y centrifugando 30 seg. Tras ello se lavaron una segunda vez con 250 μL del mismo tampón y se centrifugaron durante 2 min para secarlas completamente. La elución se llevó a cabo añadiendo 50 μL de agua directamente sobre el centro de las membranas, incubando 1 min a temperatura ambiente y centrifugando 30 seg a máxima potencia.

Los eluidos se cuantificaron con NanoDrop y se comprobaron en un gel de electroforesis en agarosa al 1.8%.

3.10 Sonicación de DNA

Para hibridar un DNA genómico de referencia de *S. cerevisiae* en los microchips de embaldosado (ver apartado *Hibridación de microchips de embaldosado*) fue necesario un primer paso de fragmentación de las cadenas de DNA hasta alcanzar un tamaño promedio de 150-200 pb. Para ello se utilizó el equipo de sonicación S-Series SonoLAB Single (Covaris). Siguiendo las instrucciones del equipo, para obtener unos fragmentos del tamaño de interés hubo que ajustar los parámetros de sonicación de la siguiente manera: 10% de ciclos de trabajo, nivel 5 de intensidad, 200 ciclos por pulso, para un total de 180 seg de sonicación. Se sonicaron 5 μg de gDNA por muestra en un volumen de 120 μL de agua, utilizando para ello microtubos. Los productos de fragmentación se cuantificaron y se comprobaron mediante electroforesis en gel de agarosa.

4. Técnicas genómicas

4.1 Genomic run-on radiactivo (GRO)

El experimento modelo para la detección de eventos de transcripción naciente en levadura a escala global es el GRO, descrito en (García-Martínez *et al.*, 2004; modificado en García Martínez *et al.*, 2011). En esta tesis se ha modificado el protocolo para estudiar la influencia de la incubación con RNasa A en la eficiencia de incorporación de nucleótidos por parte de las RNAPs nucleares. La incubación con RNasa A se usó para aumentar la resolución del BioGRO y provoca tanto la depleción de los RNAs maduros de la célula, como el *recortado* de los RNAs en síntesis unidos físicamente a las RNAPs.

Por cada experimento se emplearon 100 mL de un cultivo de levadura creciendo en fase exponencial (alrededor de una DO_{600} de 0.5). Las células se recogieron por centrifugación y, tras eliminar el medio de cultivo, se congelaron en nitrógeno líquido. Se transfirieron a -20°C y se mantuvieron a esa temperatura durante al menos 3 h. Transcurrido ese tiempo, las células se descongelaron en hielo y se permeabilizaron lavándolas con 10 mL del detergente sarkosil al 0.5%. Tras el lavado, las células se dividieron en dos alícuotas: la primera (muestra control) se mantuvo en hielo mientras la segunda (muestra digerida) fue tratada con RNasa. Para ello las células se incubaron en agitación durante 10 min a 30°C con 32 μL de RNasa A (10 mg/mL) disuelta en 3.2 mL de sarkosil. Tras la digestión con RNasa, las células fueron lavadas con tres pases de 50 mL de sarkosil para eliminar la RNasa, y se transfirieron a un nuevo tubo, donde se procedió a dar el pulso de transcripción con un nucleótido radiactivo. Las células de la muestra control fueron resuspendidas en 120 μL de agua mientras que las que habían sido tratadas con RNasa se resuspendieron en 115 μL de agua, y se les añadió 5 μL de RNase OUT para proteger los RNAs nacientes de posibles residuos de RNasas. Seguidamente se añadieron a ambas muestras 120 μL de tampón de transcripción 2.5X (50 mM Tris-HCl pH 7.7, 500 mM KCl, 80 mM MgCl_2), 16 μL de NTPs (ATP, CTP y GTP 10 mM de cada), 6 μL DTT 0.1 M y 16 μL α - ^{33}P -UTP (10 $\mu\text{Ci}/\mu\text{L}$). En los experimentos en los que se utilizó otro derivado de nucleótido

marcado (Bromo-UTP, Metil-CTP) éste se añadió a la concentración adecuada en cada caso en sustitución del α -³³P-UTP. La reacción de run-on se produjo incubando la mezcla a 30°C en agitación durante 5 min. La reacción se detuvo con 1 mL de agua a temperatura ambiente a 4°C. Las células se mantuvieron 5 min en hielo y se recogieron por centrifugación, eliminando el sobrenadante con los nucleótidos radiactivos no incorporados. La extracción del RNA se llevó a cabo mediante el método del fenol ácido descrito anteriormente, pero realizando únicamente una precipitación con 0.1 volúmenes de LiCl 5 M y 2.5 volúmenes de EtOH 96% durante 3 horas a -20°C. El RNA se resuspendió en 300 μ L de agua y se cuantificó la cantidad de RNA mediante NanoDrop. Se utilizaron 5 μ L de una dilución 1:100 de cada muestra para medir la incorporación radiactiva, usando para ello líquido de centelleo *Cocktail 22 Normascint* (CO0135) (Scharlau) y un contador de centelleo *Tri-Carb 2810 TR Liquid Scintillation Analyzer* (PerkinElmer).

Las muestras de GRO se cargaron en un gel de agarosa desnaturalizante, se inspeccionaron visualmente mediante tinción con BrEt, y se analizaron por *Northern Blot*, tal y como se ha descrito anteriormente.

4.2 Genomic run-on no radiactivo (BioGRO)

El marcaje de transcritos nacientes con 4-Tio-UTP o Biotina-UTP a escala global se realizó de la misma manera que el descrito para el GRO radiactivo en el apartado anterior, con las siguientes modificaciones: todas las muestras fueron tratadas con RNasa, y en el mix de transcripción se sustituyeron los 16 μ L de UTP radiactivo por 20.25 μ L tanto de 4-Tio-UTP (5 μ M, TriLink) como de Bio-11-UTP (10 mM, Ambion), ajustándose los volúmenes finales de reacción con agua. Tras el pulso transcripcional, para extraer el RNA tiolado se utilizó el método del fenol ácido, y para el RNA biotilado el de la proteinasa K. Seguidamente, a los RNAs tiolados se les incorporaron moléculas de biotina mediante el proceso descrito anteriormente. En ambos casos, los RNAs marcados se seleccionaron por tamaño mediante el protocolo descrito anteriormente.

Los RNAs se corrieron en un gel de agarosa al 1.8% en TBE y se analizaron mediante la técnica de Northern Blot para detectar la presencia y la cantidad relativa de biotina de las muestras.

4.3 Filter run-on

Una manera de cuantificar la actividad de las RNAP activas con una cantidad pequeña de muestra es mediante un protocolo de *run-on* desarrollado en nuestro laboratorio (Palit y Pérez-Ortín, no publicado) llamado *Filter run-on*. La técnica ha sido utilizada en esta tesis para medir la eficiencia de incorporación radiactiva en células de levadura tras la congelación rápida en nitrógeno líquido.

La primera parte del protocolo experimental se basa en adaptar los procedimientos empleados en el GRO a una escala dieciséis veces más pequeña. Se partió por tanto de un cultivo de células creciendo en medio rico YPD en fase exponencial alrededor de 0.5 de DO_{600} y se recogieron 3.125 mL de cultivo por muestra. Cada muestra se permeabilizó con 5 mL de sarkosil 0.5%, y se transfirió a un nuevo tubo. Las células se resuspendieron en 7.2 μ L de agua y se añadió una mezcla de reacción consistente en 7.5 μ L de tampón de transcripción 2.5X, 1 μ L de rNTPs, 0.375 μ L de DTT y 1 μ L de α -³³P-UTP. El pulso de transcripción se dio incubando las mezclas a 30°C durante 5 min. Se paró la reacción con 62.5 μ L de agua atemperada a 4°C, y se mantuvieron las células en hielo hasta el siguiente paso del protocolo.

En los casos en los que el experimento requirió de una depleción previa de los mRNA celulares, las muestras fueron tratadas de forma similar a las del GRO: tras la permeabilización inicial se mezclaron las células con 2 μ L de RNasa A (10mg/mL) disuelta en 200 μ L de sarkosil durante 10 min a 30°C en agitación. Tras ello se lavaron las células con 5 mL de sarkosil tres veces, se recogieron y se transfirieron a un nuevo tubo. Las células se resuspendieron en 6.89 μ L de agua y se les añadió 0.31 μ L de RNase OUT para proteger los RNAs nacientes de posibles residuos de RNasas. La mezcla de reacción y el

procedimiento de pulso transcripcional fueron iguales que los de las muestras no digeridas con RNasa.

Para cuantificar la incorporación de nucleótidos radiactivos en el RNA nascente se utilizó un método selectivo de precipitación de ácidos nucleicos con ácido tricloroacético (TCA), basado en lo descrito en el manual de laboratorio *Molecular Cloning*, apéndice A8.25 (Sambrook y Russell, 2001). Se cortaron trozos de papel de fibra de vidrio GF 6 (Schleicher y Schuell, #10370017) y se colocaron en el fondo de viales de vidrio para contaje de centelleo. Cada muestra se dividió en tres alícuotas de 15 μ L, y fue repartida con la pipeta automática por la superficie del papel de fibra de vidrio y dejada secar en una estufa a 65°C durante al menos 30 min. Transcurrido ese tiempo, uno de los tres viales de cada muestra fue separado y no se le realizó ninguna precipitación con TCA. Se le añadieron 5 mL de líquido de centelleo y se midió en contador de centelleo. De esta manera sirvió como medida de la radiactividad *total* (entendida tanto como nucleótidos libres, como nucleótidos incorporados por las RNAPs en RNAs nascentes) en contacto con las células durante la reacción de *run-on* (muestra *Total*). Los otros dos tubos se procesaron en paralelo, a modo de réplicas técnicas, y sirvieron para cuantificar específicamente la radiactividad procedente de nucleótidos formando parte de polímeros de RNA (muestras *TCA*). La precipitación se consiguió añadiendo 4 mL de TCA 10% frío por vial, y dejándolo durante 25 min a 4°C. Seguidamente se decantó el TCA, se reemplazó por otros 4 mL de TCA frío, y se volvió a dejar a 4°C 10 min. Tras ello se lavaron los viales dos veces: primero añadiendo 5 mL de EtOH 70% frío y manteniéndolos 10 min a 4°C, y después añadiendo 5 mL de EtOH 96% frío y dejándolos otros 10 min a 4°C. Tras el último lavado, el alcohol se decantó y los tubos se dejaron secar en la estufa durante al menos 30 min. A cada vial se le añadieron 5 mL de líquido de centelleo y se midió su radiactividad en el contador. Los cálculos de porcentaje de incorporación se realizaron siguiendo la relación:

$$\text{Incorporación(\%)} = \frac{dpmsTCA}{dpmsTotal} \times 100$$

4.4 Hibridación de macrochips de DNA de nailon

Para las hibridaciones de macrochips se utilizaron los fabricados por la Sección de Chips de DNA-S.C.S.I.E de la Universitat de València. Los filtros utilizados son membranas de nailon cargadas positivamente *Nytran SPC* (Whatman) sobre las que se depositaron 6144 sondas de DNA bicatenarias generadas mediante PCR utilizando el robot de fabricación de chips BioGrid (BioRobotics). Para una descripción más detallada ver Alberola *et al.*, (2004). Los filtros utilizados correspondieron a la versión 10. La hibridación de la muestra radiactiva se realizó tras pre-hibridar los filtros en Solución de Hibridación (tampón fosfato 0.5 M, EDTA 1mM, SDS 7%, pH 7.2) durante 1 hora a 65°C. Tras ello se añadió la muestra a una concentración de 5 millones de dpm/mL en un volumen de 3 mL y se dejó hibridar en horno rotatorio a 65°C durante 48 horas. Tras ello se recuperó la muestra para posibles futuros usos y se lavaron las membranas durante 10 minutos a 65°C con un pase de Solución de Lavado 1 (1X SSC, 0.1% SDS) y dos pases de Solución de Lavado 2 (0.5X SSC, 0.1% SDS). Las membranas lavadas se sellaron entre láminas de plástico y se expusieron sobre pantallas IP durante un tiempo variable, determinado por la intensidad de señal global de hibridación de cada membrana medida por contador *Geiger*. Tras adquirir la imagen los filtros fueron tratados con dos pases de Solución de *Stripping* (tampón fosfato sódico 5 mM, pH 7, SDS 0.1%) en ebullición, con el objetivo de eliminar los RNAs hibridados en los filtros y poder así reutilizarlos.

4.5 Hibridación de microchips de embaldosado

De cada muestra de BioGRO se realizaron un mínimo de dos réplicas biológicas, que se hibridaron en microchips de embaldosado *Tiling Array Custom* (PN 520055, Affymetrix, Santa Clara, CA) diseñados por Lars Steinmetz del Laboratorio Europeo de Biología Molecular (EMBL Heidelberg, Alemania), (David *et al.*, 2006). Con un diseño basado en la secuencia de referencia de la cepa de laboratorio S288c, se trata de un array de oligonucleótidos que cubre las dos cadenas de todo el genoma de *S. cerevisiae* con 6.5 millones de sondas de 25 nucleótidos dispuestas de manera que existe un solapamiento

de secuencia entre ellas de 17 nucleótidos (<http://www.ebi.ac.uk/arrayexpress/files/A-AFFY-116>). Las hibridaciones se llevaron a cabo en el Servicio de Análisis Multigénico de la Unidad Central de Investigación UCIM / INCLIVA de la Universitat de València. Se hibridaron 10 µg de RNA de BioGRO directamente sobre el chip, siguiendo los procedimientos descritos a partir del apartado 5 (*Hybridization*) del manual *GeneChip® Whole Transcript (WT) Sense Target Labeling Assay Manual*, usando el *GeneChip® Hybridization, Wash and Stain Kit* (P/N 900720).

Debido a la baja intensidad de señal quimioluminiscente que proporcionó el método directo de hibridación utilizado, se procedió a aumentar la señal mediante un pase adicional de tinción de los chips (detallado en el capítulo 6 del mismo manual).

4.6 Amplificación de la muestra de BioGRO

Se siguió el protocolo descrito en el kit de preparación de librerías de secuenciación masiva *NEBNext Small RNA Library Prep Set For Illumina* (New England Biolabs). La muestra de BioGRO se acompañó en paralelo de un control negativo de biotilación, que consistió en otra muestra procesada de la misma manera, pero con la única diferencia de que en la reacción de run-on se sustituyó el Bio-11-UTP por UTP no marcado (muestra *Mock*).

1^{er} paso: ligación del adaptador 3'. Se disolvieron 5 µg de los RNAs de BioGRO en 6 µL de agua y se les añadió 1 µL de *Multiplex 3' SR Adaptor*. La mezcla se calentó durante 2 min a 70°C en un termociclador precalentado y se dejó enfriar en hielo. Seguidamente se añadieron 10 µL de *3' Ligation Reaction Buffer 2X* y 3 µL de *3' Ligation Enzyme Mix*, para un volumen total de reacción de 20 µL. La mezcla se incubó en termociclador durante 1 hora a 25°C, y después 15 min a 65°C para inactivar la enzima.

2^o paso: refosforilación de extremos 5'. Dado que el corte endonucleolítico mediado por las RNAsas durante el proceso de BioGRO elimina los extremos 5' fosfato, generando fragmentos de RNA con extremos 5'-OH, para poder realizar reacciones de ligación sobre ellos es necesario refosforilarlos. Para ello se añadieron a los 20 µL de mezcla del paso

anterior 1 μL (10 unidades) de polinucleótido quinasa (PNK, Roche) y 2 μL del tampón PNK 10X suministrado con la enzima. Se llevó la mezcla a 30 μL finales con agua y se incubó 30 min a 37°C, manteniendo posteriormente las muestras a 4°C hasta el siguiente paso.

3^{er} paso: hibridación del primer para RT. Se añadió 1 μL de *Multiplex SR RT Primer* directamente a la mezcla del paso anterior y se incubó en termociclador 5 min a 75°C, 15 min a 37°C y 10 min a 25°C. Tras la incubación se precipitaron los RNAs durante toda la noche a -20°C añadiendo 2.5 μL de NaCl 5M, 0.5 μL de glucógeno y 62.5 μL de EtOH 96%. Al día siguiente los RNAs precipitados y lavados con EtOH 70% se resuspendieron en 25.5 μL de agua.

4^o paso: ligación del adaptador 5'. A los 25.5 μL de RNAs del paso anterior se les añadieron 2.5 μL de *5' Ligase Enzyme Mix*, 1 μL de *5' Ligation Reaction Buffer 10X* y 1 μL de 5' SR Adaptor, que había sido previamente desnaturalizado a 70°C durante 2 min y mantenido en hielo hasta su utilización. La reacción de ligación, en un volumen final de 30 μL , se produjo a 25°C durante 1 hora.

5^o paso: RT. Los RNAs unidos a adaptadores de secuencia conocida a ambos extremos se sometieron a un ciclo de transcripción reversa para convertirlos en cDNA. Para ello se tomaron 24 μL de los RNAs del paso anterior y se añadieron los siguientes componentes del kit: 2 μL de dNTPs, 8 μL de *First Strand Buffer 5X*, 4 μL de 0.1 M DTT, 1 μL de RNase OUT y 1 μL de transcriptasa reversa *SuperScript II RT*. Los 40 μL de reacción final fueron incubados durante 1 hora a 44°C y después 15 min a 70°C para inactivar la enzima.

6^o paso: purificación del híbrido RNA/DNA con Dynabeads. Para poder seleccionar aquellos híbridos RNA/DNA procedentes de la RT de un RNA naciente biotinilado, cada muestra se enriqueció con 50 μL de esferas magnéticas de estreptavidina. (tal y como se describe en el apartado *Purificación de RNA biotinilado con bolas magnéticas*).

7^o paso: PCR. Los eluidos del paso anterior sirvieron como molde para la amplificación de la muestra por PCR. La mezcla de reacción constó de 21.25 μL de eluidos, 2.5 μL de *LongAmp Taq 2X Master Mix*, 2.5 μL de *Index 1 Primer for Illumina* y 1.25 μL del oligo GX2-T7. En paralelo se amplificó un control negativo de PCR, en el que en vez de muestra se puso agua. El perfil térmico consistió en un primer paso de desnaturalización a 94°C

durante 30 seg, seguido de 16 ciclos de una sucesión de etapas de desnaturalización a 94°C durante 15 seg, anillamiento de los cebadores a 62°C durante 30 seg, y extensión a 70°C durante 15 seg. El último paso consistió en una extensión final de las cadenas sintetizadas a 70°C durante 5 min. Tras la PCR se comprobó el resultado mediante electroforesis en gel de agarosa.

8º paso: generación de cRNA mediante IVT. Para preservar la orientación de las hebras se volvió a generar el RNA original a partir del producto de PCR mediante IVT. Como paso previo se limpiaron los productos de PCR de nucleótidos no incorporados y de otros restos de mix de reacción con el kit de purificación *GeneJet PCR Purification Kit from Fermentas* (Thermo), siguiendo las instrucciones del manual. Una vez purificados, se utilizaron 10 µL de los productos de PCR de la muestra de BioGRO y de la muestra *Mock*, y se procedió de la misma manera a como se describe en apartados anteriores. Los cRNAs, libres del DNA molde de PCR se comprobaron mediante electroforesis.

5. Métodos de análisis bioinformático de los datos

5.1 Análisis de imagen de los microchips de embaldosado

Los archivos en formato .CEL con las imágenes adquiridas mediante el *scanner* de microchips se procesaron con el programa *Tiling Analysis Software* (TAS, Affymetrix), con los parámetros de detección de señal ajustados por defecto.

Para la inspección visual de las señales de hibridación en relación a las anotaciones del mapa genómico de referencia de *S. cerevisiae* se utilizó el programa *Integrated Genome Browser* (<http://bioviz.org/igb.html>).

Adicionalmente, se usó software basado en *scripts* de código en lenguaje estadístico R (<http://www.r-project.org/>) para la lectura y procesado posterior de los archivos .CEL. Entre ellos, se usaron diversos *scripts* cuyo código fue escrito por Vicent Pelechano,

Ishaan Gupta y Zhenyu Xu (EMBL Heidelberg), modificados del paquete de Bioconductor *TilingArray* (<http://www.bioconductor.org/packages/2.11/bioc/html/tilingArray.html>) y adaptados posteriormente para el desarrollo de esta tesis. Adicionalmente se utilizó el paquete de *scripts* de R *Simple tiling array analysis of Affymetrix ChIP-chip data* (Starr, Zacher *et al.*, 2011)

5.2 Normalización de los datos

5.2.1 Normalización de los datos mediante el software de Affymetrix

Para obtener datos significativos de señal de hibridación correspondientes a la muestra y eliminar sesgos técnicos producidos por la distinta capacidad de hibridación intrínseca de cada sonda del chip se normalizaron los datos comparándolos con la señal de referencia proporcionada por un gDNA procesado e hibridado bajo las condiciones estándar de Affymetrix. Para ello se utilizó el método de normalización por cuantiles.

El mismo método de cuantiles se utilizó para comparar entre sí las distintas hibridaciones de cada muestra y así sacar el promedio de las señales, sin la intervención de una hibridación de gDNA como referencia.

5.2.2 Normalización de los datos mediante el lenguaje estadístico R

Con los *scripts* de R se normalizaron los datos frente al gDNA de referencia mediante el método *Variance Stabilization and Normalization*, VSN.

5.3 Bases de datos y otras herramientas para el análisis bioinformático

Se utilizaron para el análisis las siguientes herramientas *online*:

- Base de datos SGD. "Saccharomyces Genome Database" (<http://www.yeastgenome.org/>) (Cherry *et al.*, 2012).
- La herramienta de minería de información bibliográfica *Textpresso for yeast* (<http://textpresso.yeastgenome.org/>) (Muller *et al.*, 2004).
- La potente herramienta online de análisis global del genoma de *Saccharomyces cerevisiae* *Yeast Mine* (<http://yeastmine.yeastgenome.org/yeastmine/begin.do>), que permite, a partir de listas de genes de interés, hacer búsquedas de categorías funcionales basadas en *Gene Ontology*.
- El programa para realizar cálculos masivos de diferencias por test-t: *GraphPad* (<http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=C>)
- Generador online de diagramas de Venn (<http://www.pangloss.com/seidel/Protocols/venn.cgi>)
- La base de datos de posicionamiento de nucleosomas en *Saccharomyces cerevisiae* (http://atlas.bx.psu.edu/cj/nucl_retrieval.html). (Jiang y Pugh, 2009).
- Las bases de datos de experimentos de *microarrays*, *ArrayExpress* (<http://www.ebi.ac.uk/arrayexpress>) y *GEO* (<http://www.ncbi.nlm.nih.gov/geo/>)
- La base de datos de tRNAs, *Transfer RNA Database* (<http://trna.bioinf.uni-leipzig.de/DataOutput/>).
- El repositorio de paquetes de programas basados en lenguaje R para el análisis de datos biológicos *Bioconductor* (<http://www.bioconductor.org/>)
- La interfaz de usuario para el manejo de entorno de programación basado en el lenguaje R de código abierto, *R Studio* (<http://www.rstudio.com/>).

RESULTADOS

Capítulo 1. Aspectos técnicos del método BioGRO

1.1-Antecedentes de trabajo

El presente trabajo de tesis es una continuación del trabajo previo realizados en nuestro laboratorio por Vicent Pelechano. Uno de los objetivos de su tesis (en el que se obtuvo sólo resultados preliminares) fue desarrollar una técnica de GRO de alta resolución para la medida de transcripción naciente a nivel genómico usando la plataforma de microchips de embaldosado de Affymetrix. Su trabajo en este sentido se centró en testar la incorporación al RNA naciente de nucleótidos no radiactivos, como el Tio-UTP y la Biotina-UTP, compatibles con la mencionada plataforma.

El Tio-UTP es un análogo de la uridina que es incorporado de manera eficiente por las RNAPs (Melvin, 1978), y que puede ser covalentemente unido a una molécula de biotina mediante tratamiento con un compuesto (Biotin-HPDP) que genera puentes disulfuro entre la biotina y el residuo de uridina. La biotinilación de los nucleótidos con grupos tiol incorporados *in vivo* al RNA naciente se utiliza como estrategia para la posterior separación específica del RNA maduro no biotinilado (Dolken, 2008). Tras comprobar que la incorporación de Tio-UTP por las RNAPs de la levadura durante el run-on era efectiva, los controles que utilizó para el proceso de biotinilación revelaron un artefacto muy evidente, pero de difícil explicación: la aparente incorporación no mediada por run-on de tio-uridinas a los RNAs maduros. Tal artefacto invalidaba la aproximación, debido a que no quedaba garantizado el aislamiento específico de los RNAs nacientes para su estudio.

El siguiente nucleótido que empleó fue uno que directamente lleva la biotina incorporada, la Biotina-11-UTP. Tras realizar un run-on en presencia de este análogo, y comprobar que no aparecía el mismo artefacto que con la incorporación del Tio-UTP, se decidió probar la hibridación directa de la muestra en los chips de Affymetrix. Paralelamente, se hibridó una alícuota de misma muestra procesada de tal manera que los RNAs biotinilados se separaron del resto con esferas magnéticas recubiertas de estreptavidina, y se amplificaron mediante transcripción inversa. En cuanto a la hibridación directa, los

resultados no fueron muy concluyentes debido a la débil señal obtenida de los chips, situada en el rango inferior de sensibilidad de la técnica, que sólo permitió detectar unos pocos genes de alta transcripción (figura A). Aunque la hibridación de la muestra purificada y amplificada aumentó el nivel de señal, el análisis preliminar tampoco fue muy concluyente. La comparación de los perfiles de transcripción naciente con los de RNA maduro resultaron ser muy similares, no pudiéndose confirmar categóricamente que las señales se debieran a la hibridación de uno u otro. Como perspectiva final, en la tesis de V. Pelechano se sugirieron dos estrategias distintas para mejorar el método: 1) utilizar controles específicos para confirmar que la amplificación que se realizaba era de RNAs nacientes, y no de maduros, y 2) la sustitución de la Biotina-11-UTP por otros nucleótidos de similares características como el Br-UTP o el Metil-CTP.

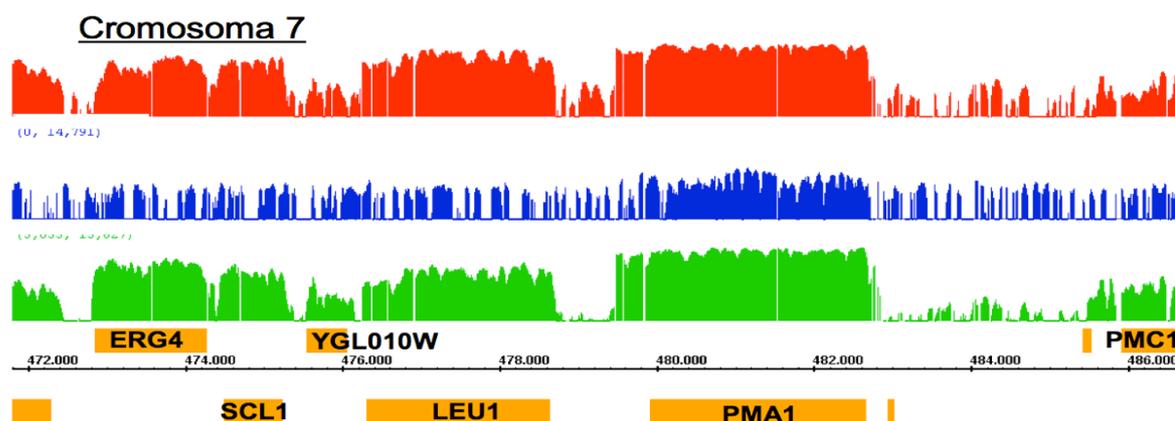


Figura A. Resultados preliminares de un GRO no radiactivo. Mapa de transcripción de una región del cromosoma VII en el que se muestran las señales de un GRO con biotina amplificado (rojo), hibridado directamente (azul), y la hibridación de un RNA maduro troceado (verde). Adaptada de Vicent Pelechano (Tesis Doctoral, 2009).

Por tanto, el presente trabajo de tesis tiene como objetivo principal la continuación de los experimentos preliminares llevados a cabo en nuestro laboratorio encaminados al desarrollo de una técnica de run-on no radiactivo de alta resolución para la medida de tasas de transcripción (TR) nacientes en *S. cerevisiae*.

1.2-Depleción de mRNAs mediante RNasa A

Caracterizar la transcripción naciente es una tarea compleja, ya que se estima que los transcritos en síntesis activa representan una proporción de únicamente el 0.34% del RNA total de una célula de levadura (Churchman y Weissman, 2011). Esta proporción tan baja de RNA naciente hace que la probabilidad de capturar RNAs maduros no deseados junto con los RNAs nacientes de interés sea muy elevada. Las técnicas desarrolladas hasta la fecha para tal fin cuentan con diversos métodos para enriquecer la muestra en RNAs nacientes de manera específica, pero ninguna está exenta de arrastrar RNA maduro contaminante, con lo que se introduce ruido en los resultados. Una manera simple de enriquecer la proporción de RNAs nacientes en la muestra de partida consiste en degradar los RNAs celulares mediante RNasa A, antes de la reacción de marcaje *in vivo* de los transcritos en síntesis mediante run-on.

El tratamiento con RNasa ha sido empleado previamente en otros trabajos para medir la extensión del run-on (Jackson *et al.*, 1998, Core *et al.*, 2008). En ambos casos se emplean núcleos permeabilizados de células humanas que se tratan con RNasa con el objetivo de recortar las partes de los RNAs más alejadas del centro activo de la RNAP, de manera que al reanudar la elongación se puede calcular la extensión de la elongación en condiciones de run-on con nucleótidos modificados. De estos estudios se extrae la conclusión de que las RNAPs elongantes protegen de la digestión una porción de ~25 bases del transcrito naciente, y que la eficiencia de incorporación de nucleótidos por las RNAPs tras el tratamiento con RNasas no se ve prácticamente afectada (Jackson *et al.*, 1998). Asimismo, se muestra que la extensión de la elongación depende de la concentración de los nucleótidos presentes durante la reacción, siendo posible controlarla limitando la disponibilidad de uno de los nucleótidos (Core *et al.*, 2008).

En el presente trabajo se adaptó la digestión con RNasa en células de levadura, empleándose como paso previo a las reacciones de run-on. Las ventajas de este proceso son: 1) se enriquece la muestra inicial en RNAs nacientes; 2) al eliminar casi completamente todos los RNAs maduros, y recortar las colas 5' de los RNAs nacientes hasta la zona protegida por la RNAP, se puede estimar la extensión de la elongación y su

dependencia con la concentración de nucleótidos añadidos para ajustar las condiciones de reacción, y por esta misma razón, 3) se puede aumentar la resolución del posicionamiento de RNAPs activas sobre el genoma de la levadura eliminando las porciones distales de los RNAs nacientes que son productos de transcripción previos a la reacción de run-on.

La primera aproximación consistió en probar si era posible deplecionar los RNAs maduros de las células de levadura en condiciones que permitieran un procesamiento posterior de run-on, así como evaluar el efecto general de una degradación nucleolítica sobre el perfil general de separación electroforética de RNAs por tamaño. Para ello se permeabilizaron las células de levadura con el detergente sarkosil y se trataron con RNasa, tras lo cual se lavaron extensivamente y se extrajo el RNA. Los RNAs extraídos se resolvieron tanto en un gel de agarosa, como en un sistema de electroforesis capilar, que permite una mayor resolución. Los resultados (Figura 1) muestran un perfil de separación por tamaño de los RNAs, en el que han desaparecido en su práctica totalidad las especies mayoritarias de RNAs maduros: las bandas de los rRNAs 25S y 18S, y las bandas correspondientes a los RNAs de pequeño tamaño, rRNA 5S, 5.8S y los tRNAs. En su lugar se encuentra una mancha difusa de acumulación de fragmentos con pico máximo en un tamaño de ~150 bases.

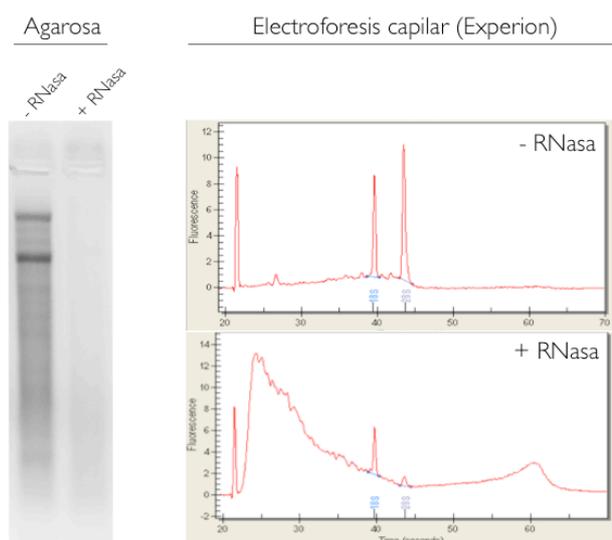


Figura 1. Perfil electroforético de RNAs de levadura antes y después de tratamiento con RNasa A. La imagen de la izquierda corresponde a un gel nativo de agarosa 2%, y la de la derecha son electroferogramas obtenidos mediante electroforesis capilar de las mismas muestras.

Para comprobar que la degradación por RNasa no impide la posterior reacción de incorporación de nucleótidos *in vivo*, procesamos diversas alícuotas de células para GRO radiactivo bajo diferentes condiciones control y experimentales. Para ello se siguió el esquema general descrito en la sección de Materiales y Métodos (*Genomic run-on* radiactivo). 50 mL de células creciendo en fase exponencial se permeabilizaron con sarkosil. Tras ello se añadió la RNasa A y se incubaron las células durante 10 min a 30°C en agitación para prevenir la deposición de las células en el fondo del tubo. Tras ello se eliminó la RNasa lavando extensivamente las células en sarkosil, y se añadió inhibidor de RNasas durante la reacción de run-on para prevenir la degradación de los RNAs sintetizados a partir de ese momento. Como controles se utilizaron las siguientes muestras: 1) una alícuota se permeabilizó con sarkosil pero no se trató con RNasa; 2) otra se permeabilizó, se trató con RNasa, pero no se lavó posteriormente para eliminar la nucleasa; 3) otra no se permeabilizó con sarkosil y se trató con RNasa y 4) una última a la que se le hizo la reacción de run-on sin añadir nucleótidos en la mezcla de reacción. La muestras experimentales consistieron en tres alícuotas permeabilizadas, tratadas con RNasa y sometidas a run-on en presencia de concentraciones crecientes de nucleótido radiactivo. Tras las reacciones de run-on, los RNAs se extrajeron, se resolvieron en un gel de agarosa desnaturalizante, y se analizaron por Northern Blot. Los resultados, mostrados en la figura 2 demuestran: 1) que el sarkosil es necesario para que las RNasas entren en las células de levadura; 2) que tras la degradación con RNasa las RNAPs son capaces de incorporar el nucleótido radiactivo mediante run-on, 3) que, a su vez, la tasa incorporación es dependiente de su concentración, y 4) que eliminar la RNasa tras la digestión es necesario para preservar la integridad de los RNAs nacientes producidos.

1.3-Aumento de la eficiencia del run-on tras congelación.

Resultados previos de nuestro laboratorio (Priyanka Palit, comunicación personal) demostraron que las alícuotas de células que se recogen para los experimentos de GRO y que no van a ser procesadas en ese mismo instante pueden congelarse y mantenerse a -20°C hasta su utilización.

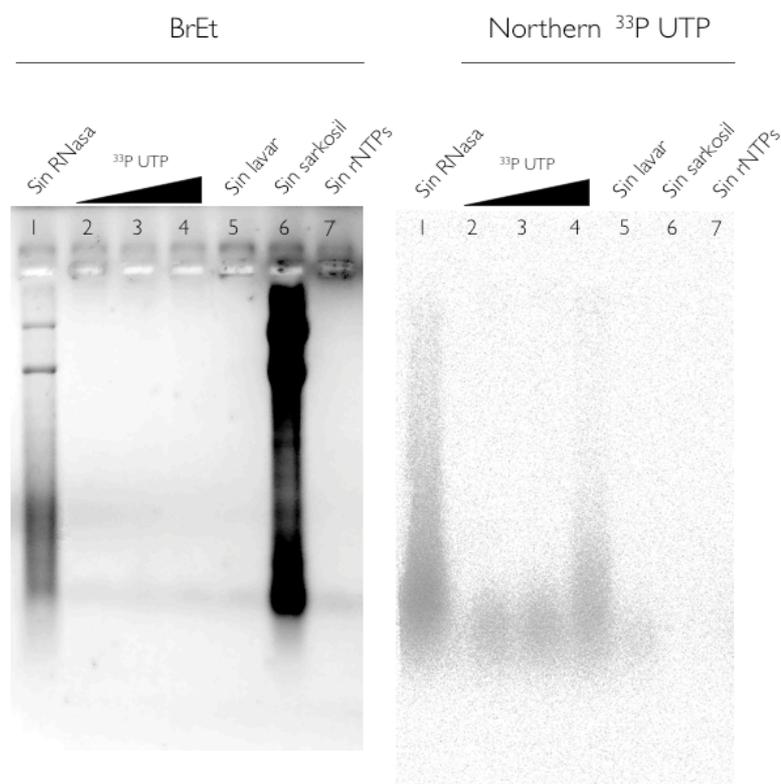


Figura 2. Incorporación de nucleótidos radiactivos mediante GRO en diferentes condiciones. Se muestra la señal de BrEt de los RNAs extraídos, y la señal radiactiva, específica de los RNAs nacientes. La muestra 1 corresponde a un GRO en el que se ha omitido el paso de degradación con RNasa. Las muestras 2-4 son GROs digeridos con RNasa y marcados en presencia de concentraciones crecientes de nucleótido radiactivo (3 μ M, 6 μ M, 12 μ M). La muestra 5 es un GRO digerido con RNasas que no han sido eliminadas. La 6 corresponde a un control negativo de GRO hecho sin haber permeabilizado previamente las células, y la 7 es un control negativo de GRO en el que se ha llevado a cabo la reacción en ausencia de nucleótidos.

También se descubrió que la congelación de las muestras debía realizarse en nitrógeno líquido, ya que recoger las células y transferirlas directamente a -20°C resultaba en una notable degradación de los RNAs celulares, debido probablemente a que el tiempo transcurrido hasta que la temperatura de la muestra baja por debajo de 0°C es lo suficientemente prolongado como para permitir la actuación de las ribonucleasas intracelulares. Una vez comprobado en el apartado anterior que es posible hacer GRO tras digestión con RNasa, quisimos testar si las condiciones de congelación interferían en

la posterior degradación por RNasas, por lo que llevamos a cabo un experimento de *Filter-GRO* (descrito en Materiales y Métodos) con dos muestras: 1) un control cuyas células no se congelaron, procediéndose directamente al ensayo de *Filter-GRO*, y 2) una muestra cuyas células se congelaron inmediatamente en nitrógeno líquido y tras dos horas a -20°C se procesaron para el ensayo. Teniendo en cuenta que los porcentajes de incorporación resultantes de un ensayo de *Filter-GRO* obtenidos en nuestro laboratorio previamente con células sin congelar se situaron en torno al 3.5% (datos no mostrados), los resultados de la Figura 3 muestran que, tal y como se había descrito previamente (Jackson *et al.*, 2008), la degradación con RNasas no disminuye significativamente el porcentaje de incorporación radiactiva, y que la congelación rápida sorprendentemente tiene como efecto la duplicación del porcentaje de incorporación. Estos datos sugieren que la congelación produce una primera rotura o debilitación de la membrana celular de la levadura, la cual se incrementa con el tratamiento posterior con el detergente sarkosil.

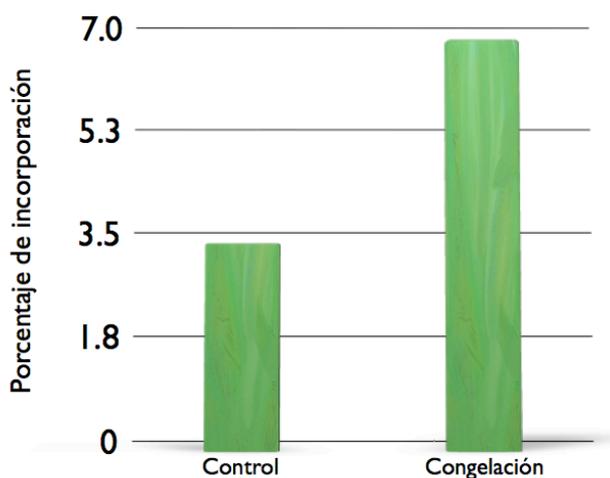


Figura 3. Aumento del porcentaje de incorporación radiactiva tras congelación. Las células se congelaron en N_2 durante unas horas. Se muestra el porcentaje de radiactividad incorporada en RNA después de 5 min. de *run-on* a 30°C .

Concluimos que este efecto combinado aumenta el grado de permeabilización de las células, posibilitando con ello unas mejores condiciones para la reacción de *run-on*. A la luz de estos resultados, en todos los ensayos de GRO hechos a partir de ese momento se

congelaron las células en nitrógeno líquido antes de la permeabilización, como estrategia para mejorar el rendimiento del marcaje de los RNAs nacientes.

1.4-Incorporación de nucleótidos modificados mediante run-on tras RNasa

Lo siguiente que nos planteamos fue aplicar las mejoras aportadas con la digestión con RNasas a las estrategias de GRO en el que se sustituyera el nucleótido radiactivo (^{33}P -UTP) por análogos no radiactivos de distinta naturaleza. Las distintas aproximaciones se describen en la figura 4.

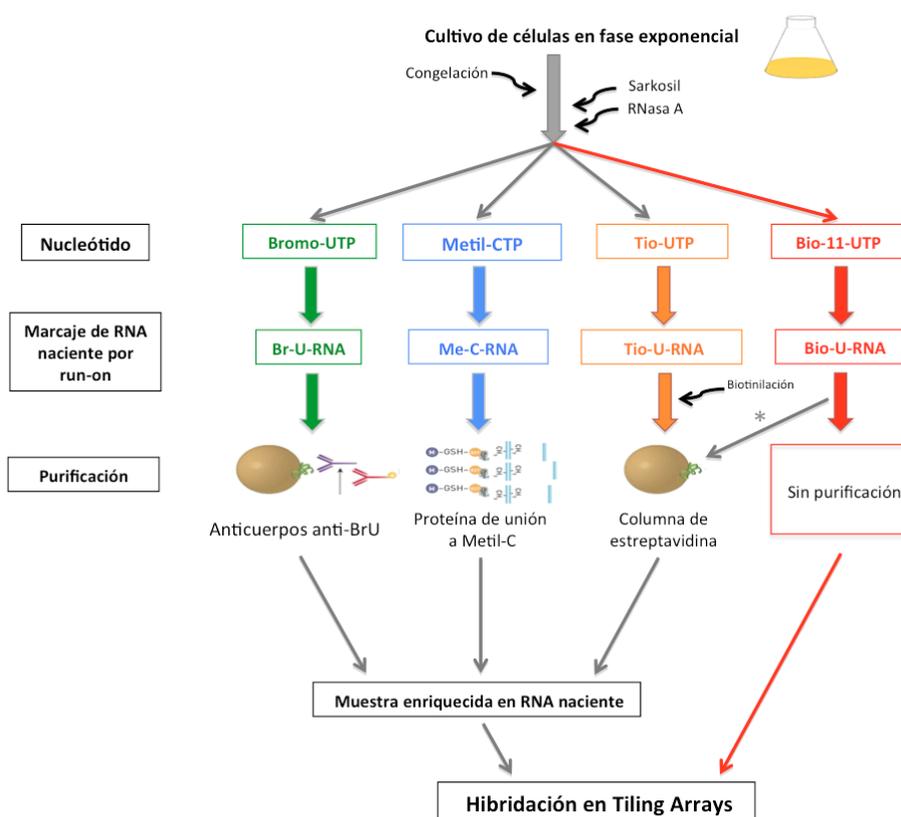


Figura 4. Esquema de las distintas aproximaciones empleadas en este trabajo para el marcaje y enriquecimiento de RNAs nacientes. * La estrategia de purificación del RNA biotinilado fue ensayada previamente (Vicent Pelechano, Tesis Doctoral. Ver apartado de Antecedentes)

Una vez marcados los RNAs *in vivo* siguiendo el esquema anterior, se utilizaron para hibridación directa (Bio-RNAs), o se aislaron selectivamente mediante varios métodos basados en las características químicas de cada tipo de análogo (figura 4).

1.4.1.-Tio-UTP

Se realizaron dos experimentos de GRO con Tio-UTP: uno sin digerir con RNasa (con el objetivo de replicar los resultados obtenidos previamente en nuestro laboratorio) y otro con digestión con RNasa. Seguidamente, para revelar la presencia de los RNAs marcados, se trataron las muestras con un reactivo biotinizante, tal y como se describe en la sección de Materiales y Métodos. Para evaluar la selectividad del proceso de biotinción también se trató en las mismas condiciones una muestra de RNA total obtenida mediante extracción estándar de ácidos nucleicos. Tanto las tres muestras anteriores, como un control negativo de RNA total sin biotinción, se separaron por tamaño en un gel de agarosa y se transfirieron a una membrana de nailon para analizar la presencia de los fragmentos biotinilados (figura 5). Los resultados del Northern (carreras 1 y 3) muestran que gran parte la señal de biotina proviene de moléculas con el mismo tamaño que los RNAs maduros de las especies mayoritarias de la levadura; es decir, rRNAs 25S, 18S, 5.8S y 5S, además del dsRNA de 4.6 kb del virus L-A (Fujimura, 2010). Esto confirmara los resultados obtenidos previamente en nuestro laboratorio, en los que se encontró que aparentemente residuos TioI se incorporan a los RNAs, de manera independiente del run-on. A pesar de los resultados, y dado que la estrategia de degradación con RNasa soluciona parcialmente el problema de la biotinción inespecífica, decidimos ir adelante con el GRO con Tio-UTP, añadiendo no obstante etapas de purificación adicionales a modo de unión de los RNAs biotinilados a esferas magnéticas recubiertas de estreptavidina. Para ello se siguió la estrategia descrita en Materiales y Métodos para la purificación selectiva de RNAs biotinilados. A modo de control de la eficiencia de purificación se utilizaron un par de RNAs sintéticos de secuencia conocida (de la bacteria *Bacillus subtilis*) generados *in vitro*, uno en presencia del nucleótido modificado, y otro con el nucleótido natural. Estos transcritos se añadieron a la muestra al principio del

proceso y se monitorizan a lo largo de las etapas de la purificación. Sin embargo, en nuestro caso no fue posible seguir este método ya que la transcripción inversa con Tio-UTP falló sistemáticamente. En su lugar se siguió el proceso de purificación de manera indirecta mediante la realización de un GRO en el que se marcaron los RNAs nacientes con una combinación de Tio-UTP y un trazador radiactivo (^{33}P -UTP, en proporciones 10:1).

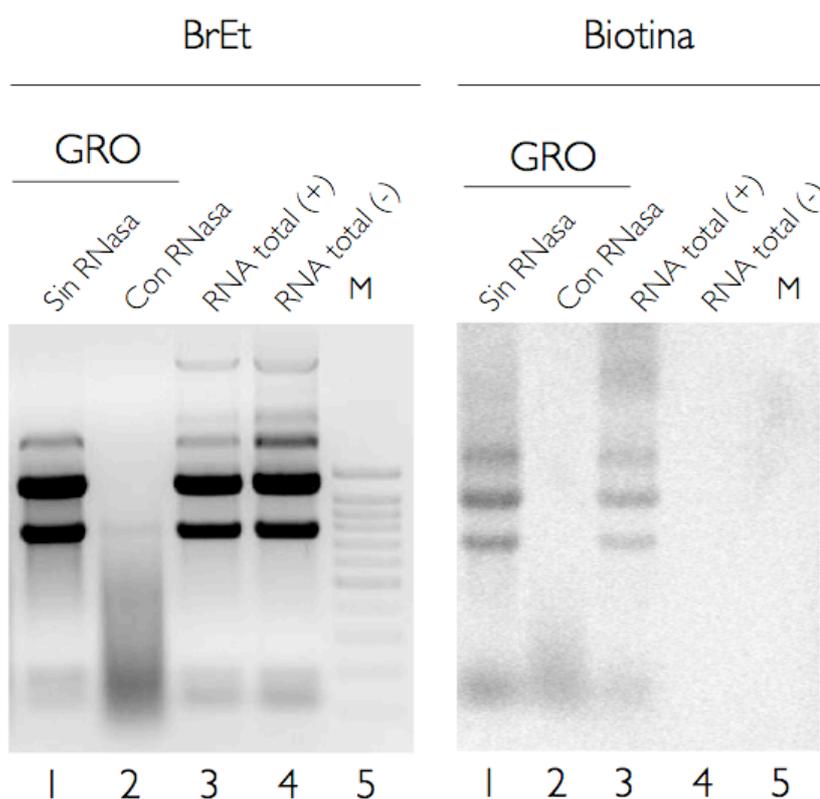


Figura 5. Selectividad de la biotilación de RNAs tiolados. Se muestra la señal obtenida por tinción con BrEt de los RNAs y la procedente de la biotina específicamente. La primera muestra es un GRO con Tio-UTP sin el paso de digestión con RNasa. La segunda es un GRO con Tio-UTP digerido con RNasa. La tercera muestra es un control de RNA total incubado con el reactivo biotilante. La cuarta muestra es un control de RNA sin incubar con reactivo biotilante. M, marcador de tamaños de múltiplos de 100pb.

Se contaron las dpm radiactivas de alícuotas de la muestra en las siguientes fracciones: inicial (antes de incubar con las esferas), no retenida (sobrenadante tras incubación con las esferas), lavados y elución. Tras repetir el proceso desde el principio varias veces se

observó que por mucho que se aumentara la cantidad de esferas magnéticas utilizadas, únicamente un 16% de la radiactividad inicial quedaba unida tras la incubación y los lavados, y que únicamente el 3% de la radiactividad inicial se eluía en el paso final. Se trató de mejorar la eficiencia de la elución aumentando la concentración de DTT, con idénticos resultados. Con estos datos concluimos que la purificación de los RNAs tiolados mediante este método es insuficiente.

1.4.2.-Bromo-UTP

Una vez descartado el uso del Tio-UTP, realizamos una búsqueda bibliográfica y encontramos que diversos autores habían hecho uso del análogo Bromo-UTP para la caracterización de la transcripción naciente en diversos organismos distintos a la levadura. Las estrategias utilizadas por los autores se basaban en uno caso en el marcaje metabólico de RNAs seguido de IP con anticuerpos anti-BrU en células de mamíferos (Jackson *et al* 1993, Ohtsu *et al* 2008), y en otro caso en el marcaje mediante *run-on*, seguido también de IP, en células humanas (Core *et al.*, 2008; para una revisión ver Friedel y Doelken, 2009). En el presente trabajo decidimos testar el marcaje con Br-UTP en levadura mediante *run-on*. Para ello preparamos primero las células para la reacción de *run-on* de la misma manera que en el apartado anterior, y realizamos el pulso de transcripción en presencia de 5-Bromo-UTP. La incorporación de Bromo-U en los RNAs nacientes se comprobó mediante *dot blot* (datos no mostrados) y se inmunoprecipitó con anticuerpos que reconocen los residuos de bromo incorporados. Dado que existían dos aproximaciones similares a la IP de RNAs con bromo descritas en la literatura, decidimos testar ambas y comparar su eficiencia. Para ello nos basamos en los procedimientos descritos en los artículos de Ohtsu *et al.*, (2008) y Core *et al.*, (2008). La diferencia fundamental entre las estrategias radica en la utilización de anticuerpos anti-BrU de dos casas comerciales distintas: mientras que el grupo de Ohtsu y col. realiza una IP con esferas magnéticas que se recubren del anticuerpo, el grupo de Core y col. utiliza esferas de agarosa pre-conjugadas con el anticuerpo (para más detalle, véase la descripción en la sección de Materiales y Métodos). En ambos casos, la manera de medir la eficiencia de la purificación consistió en monitorizar la presencia mediante qPCR de transcritos de secuencia conocida sintetizados mediante transcripción inversa en presencia, o bien de

UTP, o bien del análogo 5-Bromo-UTP. Los experimentos se llevaron a cabo un mínimo de tres veces. Tras analizar los datos de las qPCR, se estimó que el grado de enriquecimiento promedio obtenido siguiendo el procedimiento de Ohtsu y col. fue de 48 veces más RNA marcado que no marcado en la fracción IP al final del proceso. En el caso del método de Core y col., los resultados fueron de 8 veces más RNA marcado respecto del no marcado tras la purificación. Estos valores de enriquecimiento distan considerablemente de los que obtienen los autores de los trabajos, siendo de casi el doble (83 veces) el grado de purificación en el caso de Ohtsu y col, y de 1250 veces en el caso del grupo de Core y col. Por tanto, y teniendo en cuenta, como ya se ha dicho, que menos del 1% del total del RNA es de nueva síntesis, es fácil asumir que la muestra enriquecida contendrá niveles de contaminación demasiado elevados. Tras intentar optimizar la eficiencia de ambos procesos a diversos niveles y obtener similares resultados de enriquecimiento, decidimos descartar la IP de RNAs con Br-U como estrategia para caracterizar de manera fiable la transcripción naciente en levadura.

1.4.3.-Metil-CTP

Al mismo tiempo en que se realizaron los experimentos con el Br-UTP, apareció un trabajo en el que se realizaba una purificación selectiva de DNA metilado, basada en el reconocimiento de grupos metilo por parte de una proteína de fusión en C-terminal de la Glutación-S-Transferasa (GST) con el dominio de unión a grupos metilo (*Methyl Binding Domain*, MBD) de la proteína humana MeCP2p, portando una etiqueta His6 (Kangaspeska, 2008). A pesar de tratarse de un sistema diseñado y testado para el aislamiento de islotes de dinucleótidos CpG metilados en DNA, decidimos estudiar la posible eficiencia del reconocimiento de RNAs metilados. Para ello, y tras realizar la reacción de GRO en presencia de 5-Metil-CTP y comprobar la presencia de Metil-C en los RNAs nacientes mediante *dot blot* (datos no mostrados), procedimos a aislar el RNA marcado adaptando el procedimiento descrito por la autora del trabajo (descrito en Materiales y Métodos). El seguimiento de la purificación se llevó a cabo monitorizando por qPCR a lo largo del proceso la presencia de transcritos de secuencia conocida sintetizados con 5-Metil-CTP, o CTP. Tras repetir el experimento de purificación al menos tres veces, los datos de qPCR indicaron unos niveles de purificación promedio de 11 veces

más RNA marcado que no marcado en la fracción IP al final del proceso. Contando con que en la levadura la proporción inicial entre RNA maduro y naciente es de 290 veces a favor del RNA maduro, los niveles de purificación obtenidos con este método no son suficientes para el aislamiento de los RNAs nacientes con Metil-C, por lo que se descartó este tipo de marcaje.

1.5-BioGRO.

Dado que con las estrategias descritas hasta ahora de marcaje y posterior purificación selectiva de RNAs nacientes no se obtuvo el rendimiento esperado, decidimos retomar los experimentos llevados a cabo en nuestro laboratorio previamente (Vicent Pelechano, Tesis Doctoral, 2009) en relación a la hibridación directa de RNAs biotinilados mediante GRO. Esta estrategia posee la ventaja de que la señal de hibridación proviene inequívocamente de los RNAs nacientes, que son las únicas moléculas que presentan residuos de biotina, pero tiene el inconveniente de que la señal global de hibridación es baja, debido a que prácticamente todo el RNA de la muestra para hibridar en el microchip es RNA maduro no marcado, que posiblemente compite y desplaza la hibridación de los RNAs nacientes biotinilados (que constituyen una fracción muy minoritaria del total). No obstante, existen teóricamente varias maneras de minimizar esta desproporción: 1) degradar con RNasa previamente al run-on para eliminar todo lo posible los RNAs maduros de las células (de manera que para una misma cantidad de RNA hibridado en el microchip la proporción de la fracción marcada sea mayor, no tan susceptible de ser desplazada por la no marcada), y 2) visto en el apartado 1.2 que tras separar electroforéticamente los RNAs de GRO la señal de interés se concentra en una región en promedio menor de 200 bases, seleccionar los fragmentos menores o iguales a ese tamaño, descartando los restos de RNAs maduros no marcados de mayor tamaño. Con el objetivo de testar estas aproximaciones se realizó el marcaje de GRO en presencia de Bio-11-UTP (figura 6). Con el objetivo de abreviar, de ahora en adelante en esta tesis se hará referencia a esta técnica con el nombre de BioGRO.

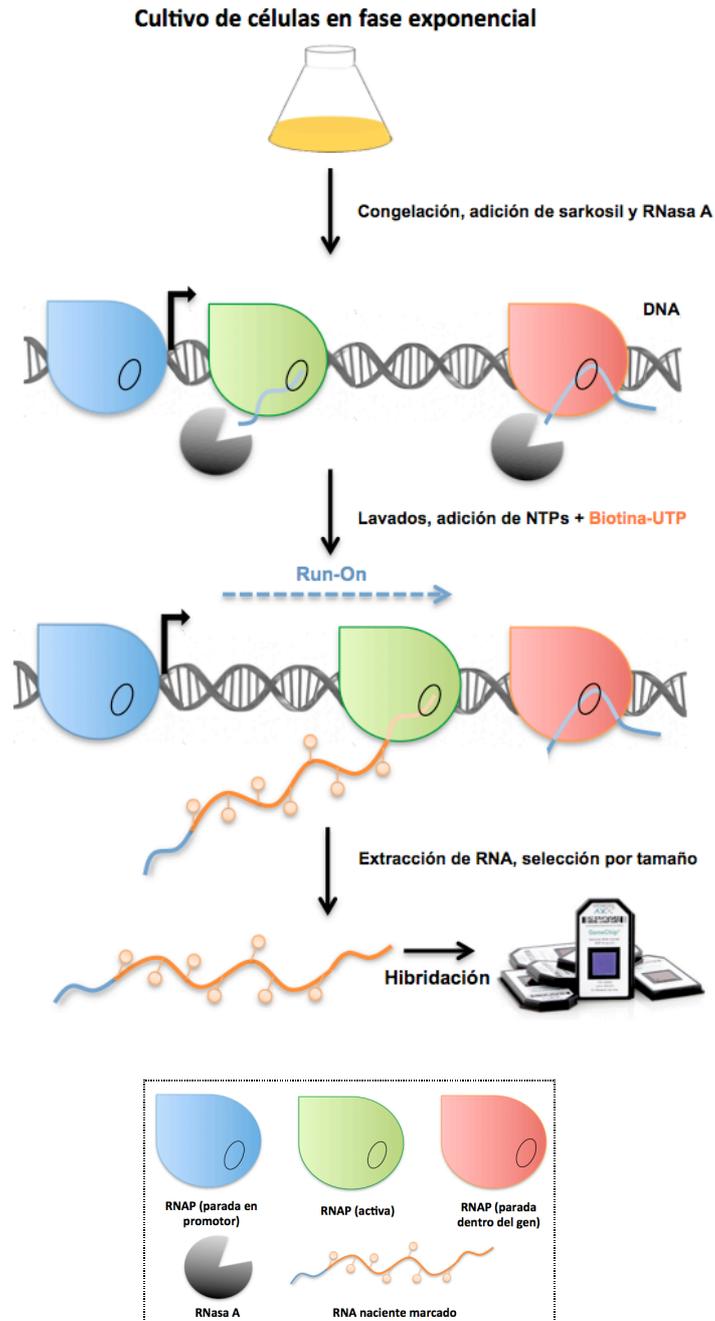


Figura 6. Esquema del método BioGRO. A partir de células recogidas en fase de crecimiento exponencial y congeladas en nitrógeno líquido, se paró la transcripción en curso añadiendo sarkosil, y los RNAs celulares se degradaron incubando con RNasa A. Tras lavar la RNasa, se dió un pulso de run-on con Biotina-UTP, se extrajo el RNA, se seleccionó por tamaño, y se hibridó en un microchip de embaldosado. Los colores de las RNAPs representan los diferentes estados transcripcionales. Sólo las RNAP activas (verde) serán capaces de hacer run-on.

1.5.1. Análisis del tamaño de elongación con biotina en distintas cepas.

El RNA se extrajo y se analizó por Northern con el objetivo de determinar el tamaño de la elongación las RNAPs con Bio-11-UTP. Para ello, y tras separar el RNA por tamaño en un gel desnaturizante de poliacrilamida (PAGE-urea), se reveló la señal de biotina con métodos quimioluminiscentes (figura 7, panel A). Por comparación con un marcador de tamaño se deduce que el rastro difuso de señal está comprendido entre las bandas de 30 y 90 bases del marcador, siendo esta última posicionada en el gel por extrapolación a partir de las bandas de mayor tamaño. A su vez, la imagen revela la existencia de un máximo de señal para los fragmentos de un tamaño alrededor de 50 bases. Asumiendo que la degradación por RNasa se ha producido de manera eficiente, y que cada una de estas moléculas posee una porción 5' de 25 bases resistente a la degradación no marcada por run-on, (Jackson *et al*, 1998), podemos concluir que las RNAPs han sido capaces de incorporar 25 nucleótidos en promedio durante la reacción de BioGRO.

Dado que uno de los objetivos de la presente tesis es aplicar el método de GRO no radiactivo al estudio de mutantes, quisimos determinar si la capacidad de elongación observada para la cepa silvestre era la misma en cepas mutantes relacionadas con la transcripción, lo cual, de no ser así, podría afectar a las conclusiones. Para testar esta hipótesis empleamos el método BioGRO en una cepa carente de una de las subunidades pequeñas de la RNAPII, la proteína Rpb4 (cepa *rpb4Δ*), y en una cepa carente de el factor de elongación TFIIS/Dst1 (cepa *dst1Δ*). Los RNAs marcados fueron separados por tamaño en geles de agarosa, y la señal de biotina se reveló de la misma manera descrita anteriormente. En el panel B de la figura 7 se observa que el tamaño de los fragmentos no es menor en el caso de los mutantes, lo que indica que las mutaciones específicas en estos factores no afectan a la incorporación de Bio-11-UTP durante el run-on.

1.5.2. Selección de RNAs por tamaño

En este estudio se ha determinado que los RNAs de BioGRO tienen un tamaño menor de 200 bases, por lo que una segunda manera de enriquecer la muestra en transcritos

nacientes es aislar las moléculas pequeñas (<200 bases) del resto para su análisis. Para ello se usó el kit de aislamiento de RNAs pequeños de la casa comercial Macherey Nagel (ver Materiales y Métodos). Se procesaron 10 µg de un RNA total de levadura siguiendo las instrucciones del fabricante y se cargaron cantidades iguales del RNA de partida y de las dos fracciones en un gel de agarosa que se tiñó con BrEt. Los resultados (panel A de la figura 8) muestran que las especies pequeñas se aislaron selectivamente de las grandes (rRNAs 18 y 25S), poniendo de manifiesto la efectividad del método.

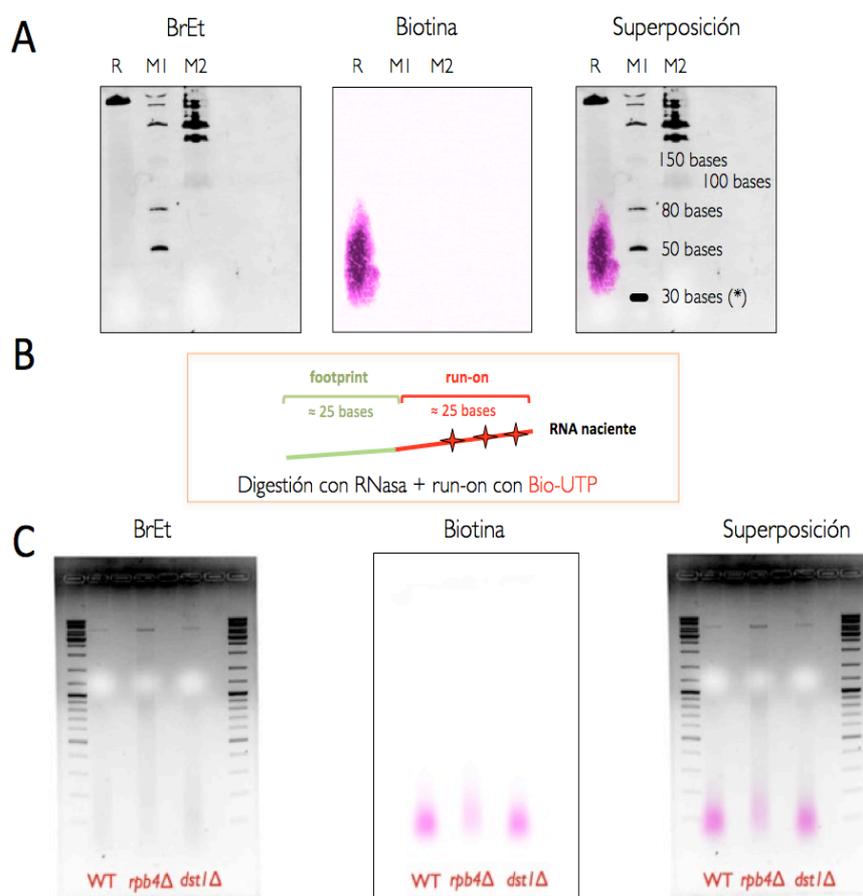


Figura 7. Análisis del tamaño de elongación de run-on con Biotina-UTP. (A) RNAs de BioGRO separados por tamaño en gel PAGE-Urea y analizados por Northern. Los RNAs biotinilados (R) aparecen como una mancha de señal de luminiscencia difusa centrada alrededor de 50 bases. * La banda de 30 bases ha sido calculada por extrapolación. **(B)** Sustrayendo la huella de la RNAP, deducimos que la elongación es de 25 bases en promedio. **(C)** Estudio del tamaño de elongación en run-on de mutantes de transcripción.

En el panel B se muestran los resultados de haber procesado de la misma manera un RNA de BioGRO. La efectividad del método en este caso se evaluó mediante Northern y revelado de la señal de biotina. Al analizar comparativamente la señal de biotina de la fracción pequeña con la del RNA de partida y la fracción grande, se vio que, a pesar de que hay una pequeña parte de la señal que se queda en la fracción grande -con lo que se pierde para el análisis posterior-, la mayoría de la señal se concentra en la fracción pequeña, por lo que se puede decir que ha habido una selección por tamaño efectiva. Complementariamente, en el gel de BrEt se muestra que una gran parte del RNA no marcado con biotina se eliminó de la muestra de interés (presente en la fracción grande y ausente de la pequeña).

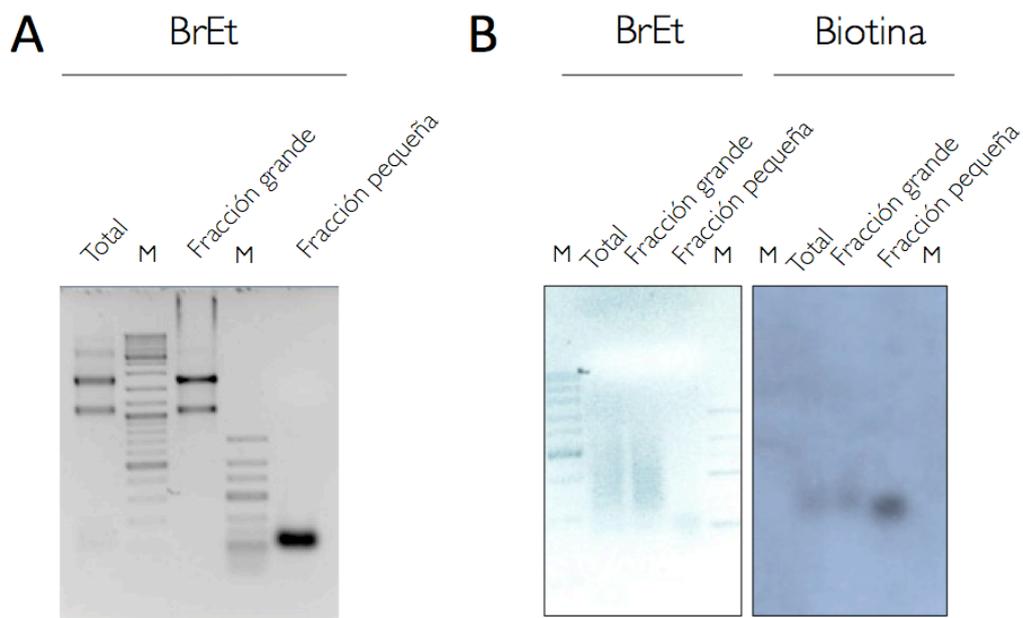


Figura 8. Selección de RNAs por tamaño. (A) Gel de agarosa en el que se han separado por tamaño alícuotas del mismo RNA total. **(B)** Gel de agarosa analizado por Northern en el que se han separado por tamaño RNAs de BioGRO. Aunque queda parte de la señal de biotina en la fracción grande, la señal aparece comparativamente enriquecida en la fracción pequeña.

1.6-Hibridación de la muestra

Se usó la plataforma de microchips de embaldosado *Tiling Array Custom* de Affymetrix (PN 520055), cuyas sondas de 25 nt cubren completamente, con un solapamiento de 8 nt las dos hebras del genoma de *S. cerevisiae* (David *et al.*, 2006). La hibridación directa de RNAs enriquecidos dio como resultado un nivel de señal mayor que en aproximaciones anteriores de nuestro grupo, pero aún así bajo, cercano al límite inferior de sensibilidad del escáner (figura 9, A).

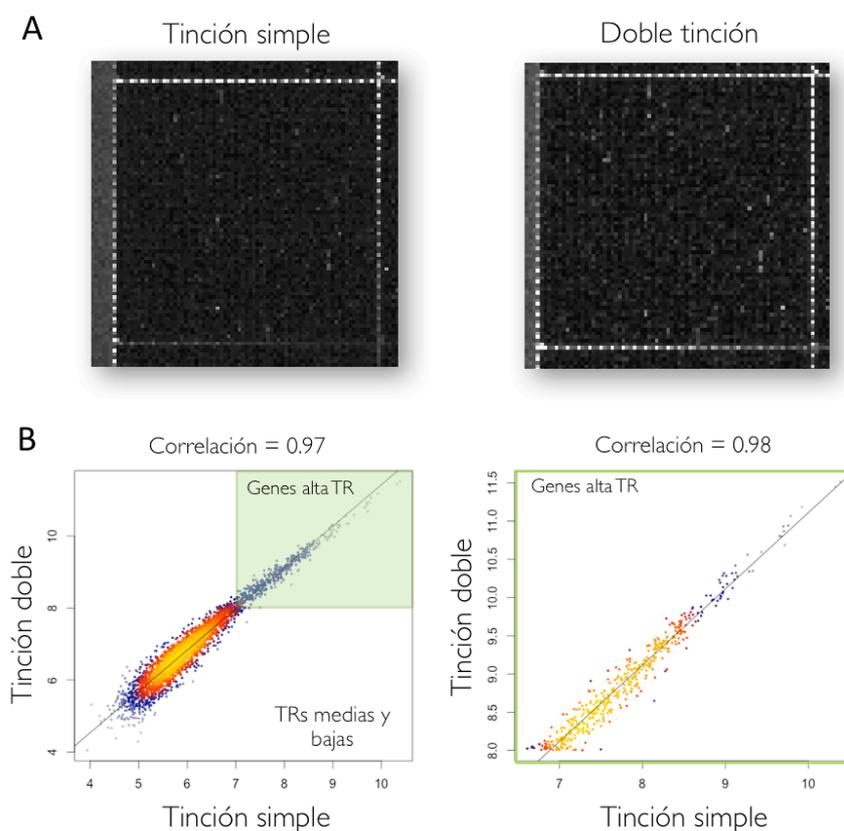


Figura 9. Estrategia de doble tinción de los *microarrays* de embaldosado. A) Tomando como ejemplo un cuadrante al azar del microchip, la señal quimioluminiscente de los *spots* aumenta visiblemente. B) Al comparar las señales de hibridación de todos los genes, se observa que la señal se ha duplicado con la doble tinción, y que ésta aumenta la señal proporcionalmente de cada spot, sin perturbarla.

Para aumentar la señal se optó por realizar una ronda adicional de tinción de los chips, modificando para ello el protocolo estándar de manejo de Affymetrix, tal y como se describe en Materiales y Métodos, bajo la supervisión del Servicio de Análisis Multigénico de la Unidad Central de Investigación UCIM / INCLIVA de la Universitat de València.

La doble tinción resultó en un aumento del doble de señal, sin artefactos o sesgos, tal y como muestra el elevado grado de correlación para la comparación entre las tinciones (figura 9, B). La triple tinción no resultó en un aumento de la señal proporcional entre genes, por lo que fue descartada.

Dada la naturaleza de la metodología de run-on *in vivo* utilizada, en la que se marcan los transcritos nacientes con un solo nucleótido modificado (en este caso, la uridina), existe la posibilidad teórica de que los resultados tuvieran un sesgo por el que las sondas más ricas en uridinas dieran más señal de fluorescencia. Para descartarlo, se hizo una comparación entre la señal de BioGRO y la composición de bases del DNA, a escala genómica. Para ello se segmentó el genoma en porciones de 100 nt, y se tomaron los valores promedio de señal de BioGRO y de porcentaje de G+C. Tras ello se tomó una muestra aleatoria de 10.000 datos que se representó gráficamente (figura 10, A). A modo de control se analizaron, de la misma manera que los datos de BioGRO, datos de posicionamiento de nucleosomas (provenientes de Kaplan *et al.*, 2009), para los que se ha descrito que existe una preferencia de posicionamiento en zonas ricas en G+C (Kaplan *et al.*, Zhang *et al.*, Tillo y Hughes, 2009). Los resultados de BioGRO ponen de manifiesto que no existe una relación, ni directa ni inversa, entre regiones ricas en A+T y la señal de hibridación. En cuanto a nucleosomas, los resultados (figura 10, B) muestran que, en efecto, existe una correlación positiva entre el porcentaje de G+C y el posicionamiento de nucleosomas (Ro de Spearman = 0.5).

Una vez obtenido el máximo grado de señal posible, se eliminaron los posibles sesgos en la capacidad de hibridación intrínseca de las sondas del chip normalizando frente a la señal de un DNA genómico troceado por sonicación e hibridado en los mismos chips mediante el protocolo estándar de Affymetrix. Elegimos el método de normalización por cuantiles (Amaratunga, 2001), implementado por defecto en el programa de análisis de microchips de embaldosado *Tiling Analysis Software*, TAS (Affymetrix).

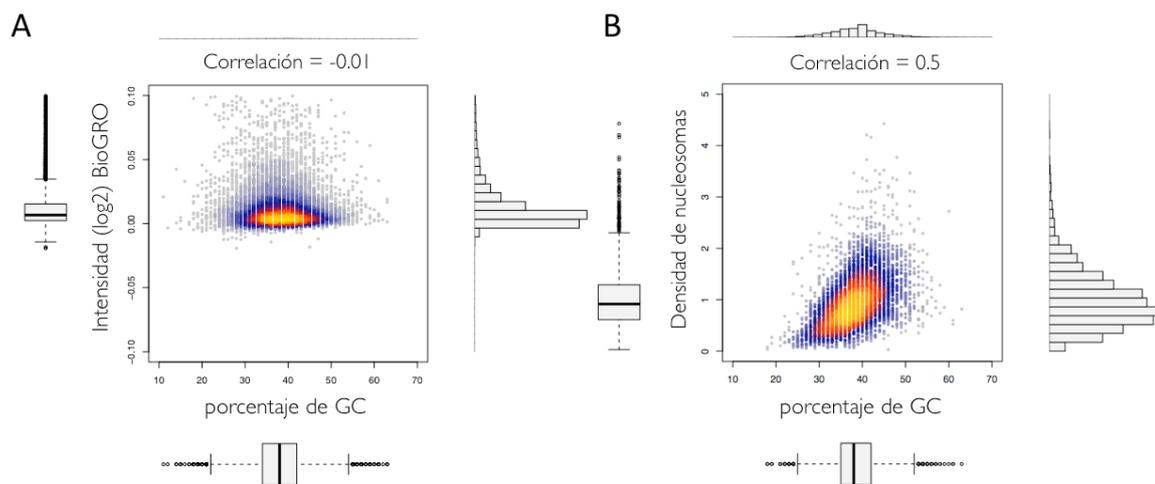


Figura 10. Análisis de la dependencia entre densidad de RNAPs activas o nucleosomas con la composición de bases del genoma.

Seguidamente, procedimos a alinear los datos normalizados de hibridación sobre el mapa genómico de la levadura, obtenido de la base de datos del genoma de *S. cerevisiae* (SGD). Una vez alineados, los inspeccionamos visualmente con el software *Integrated Genome Browser* (IGB), y tratamos de encontrar perfiles característicos del RNA naciente que nos permitieran confirmar la autenticidad de los datos. Para ello mapeamos las señales procedentes de la hebra Crick de un BioGRO y de un RNA maduro preparado mediante el método estándar de Affymetrix para chips de embaldosado. A partir de la figura 11 extraemos las siguientes diferencias entre las muestras:

- Tal y como se espera de un RNA naciente, la señal de BioGRO se extiende más allá de las anotaciones de las ORF, tanto por 5' como por 3', mientras que las señales de transcritos maduros se limitan a las zonas codificantes.
- La presencia de señal alta de RNAs nacientes en la región de los intrones de los genes *RPS6A* y *RPL19A*, específicamente en los datos de BioGRO, refleja que se trata mayoritariamente de RNA naciente marcado durante el *run-on* que no ha tenido tiempo de ser procesado por la maquinaria de corte y empalme de intrones.
- La presencia de señal en la hebra Crick en el caso del gen mapeado en la hebra Watson YPL088W sugiere la existencia de transcripción antisentido detectable

mediante el método de BioGRO. Este dato es consistente con el hecho de que los antisentido son transcritos rápidamente degradados en la célula, y por tanto difíciles de detectar en una muestra de RNA maduro.

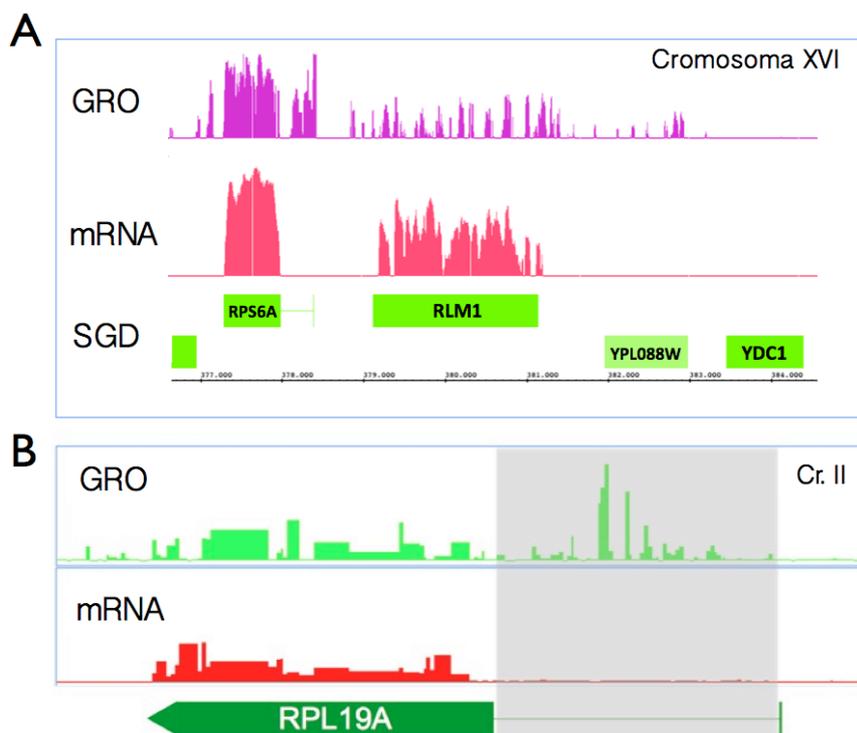


Figura 11. Distribución de RNAs nacientes sobre el mapa genómico de levadura. (A) Región del cromosoma XVI en el que se aprecia señal específica de hebra. En el gen *RLM1* la señal va más allá de las coordenadas de inicio y fin del SGD, y en el caso del gen *RPS6A* aparece señal dentro de la región intrónica, sólo en el caso de la muestra de BioGRO. **(B)** Otro ejemplo de señal dentro de la región del intrón del gen *RPL19A*, visible únicamente en la muestra de GRO.

1.7- Correlaciones de los datos de BioGRO con otros datos de TR de la literatura.

Los valores de intensidad de hibridación para cada gen se calcularon mediante el uso de programación a medida en lenguaje estadístico R (<http://www.r-project.org/>). Para ello se tomó como gen la región comprendida entre el sitio de inicio de la transcripción, TSS, y

el sitio de poliadenilación, TTS. Teniendo en cuenta de esa manera la señal procedente de las RNAP activas a lo largo de todo el transcrito, sin limitarnos a la zona codificante.

Se tomó el promedio de la intensidad normalizada de las sondas correspondientes a cada gen, y estos valores fueron transformados a escala logarítmica y comparados con otros conjuntos de datos genómicos de interés, tras estandarizarlos mediante *Z-score*. En las comparaciones mostradas en la figura 11 observamos que el índice de correlación de Pearson va desde $R=0.52$ a $R=0.67$, valores que entran dentro del rango habitual para las comparaciones de este tipo (Pelechano *et al.*, 2010). En cada caso se destacan en un color distinto al resto el conjunto de genes de proteínas ribosómicas, RP ($n=137$), dado que se trata de un grupo paradigmático de genes de alta transcripción que se comporta como un regulón transcripcional (Warner *et al.*, 1999, Pelechano *et al.*, 2009), y cuya posición relativa en la nube de puntos nos sirve para ilustrar las posibles diferencias en las medidas de TR entre las distintas metodologías con las que se obtienen. Respecto a estos genes, la comparación con los datos de SR obtenidas por el grupo de Patrick Cramer mediante el método *Dynamic Transcriptome Analysis*, DTA (Miller *et al.*, 2011) muestra una ligera tendencia a tener valores algo más altos en BioGRO (figura 12, A). Dado que el DTA mide la tasa de aparición de los RNAs en el citoplasma, este sesgo podría estar reflejando que una mayor proporción de la transcripción naciente sobre los genes RPs se transforma en moléculas maduras en el citoplasma, en relación al conjunto de los genes. La comparación con la TR derivada de la medida de densidad de RNAPII unida al DNA de estos genes (figura 12, B), obtenida por el mismo grupo de Patrick Cramer mediante IP de cromatina, CHIP (Mayer *et al.*, 2010), muestra un sesgo claro en la distribución de los genes RP hacia mayor valor en CHIP.

La técnica de CHIP mide la densidad de RNAPs unidas al DNA, sin discriminar el grado de RNAPs que se encuentran en transcripción activa, o en los diferentes estados de pausa o parada a lo largo de las regiones transcritas. Analizando los resultados, se calcula que existe un 36% más de RNAPII transcripcionalmente inactiva presente en los genes RP que en el promedio del genoma (Pérez-Ortín *et al.*, 2011), lo cual constituye una característica general de la regulación de estos genes (Pelechano *et al.*, 2009). La comparación con la metodología de GRO radiactivo de nuestro laboratorio (figura 12, C) obtiene, como era de

esperar, el grado de correlación más alto. Cabe destacar que la posición relativa de los genes RPs muestra una ligera tendencia a tener valor más alto en el BioGRO.

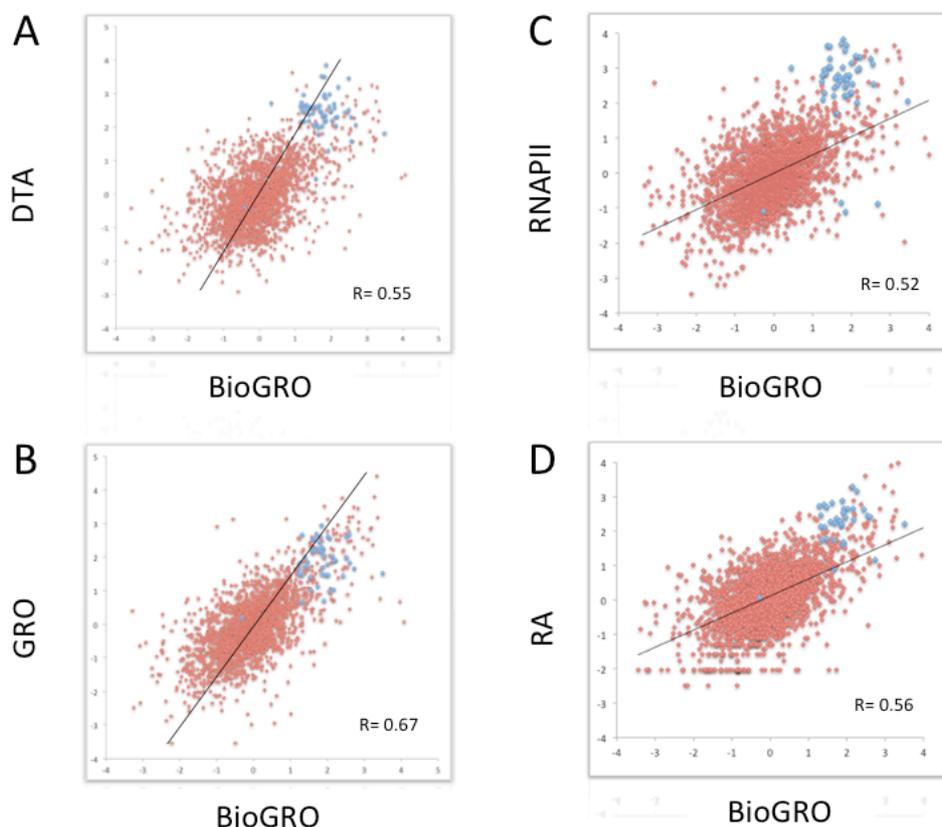


Figura 12. Correlaciones entre los datos de TR calculados por BioGRO y por otras técnicas genómicas. Los datos de TR se representan en escala logarítmica (\log_2) tras estandarización por Z-score. En la esquina derecha inferior de cada gráfico se muestra el coeficiente de *Pearson* para cada correlación. Los genes RP se representan en azul, y el general de los genes en rojo. DTA es el cálculo de SR (Miller *et al.*, 2011), RNAPII son datos de CHIP del laboratorio de Patrick Cramer (Mayer *et al.*, 2010), GRO son datos de TR naciente de nuestro laboratorio (Pelechano *et al.*, 2010), RA es la medida de cantidad de RNA mensajero (Wang *et al.*, 2002).

Esto indicaría que la técnica no radiactiva es capaz de capturar mejor la transcripción naciente en esos genes que la radiactiva, para lo cual no tenemos una explicación clara. En la comparación de los datos de BioGRO con los de concentración de mRNA maduro (RA) se observa, como era de esperar, una

correlación general menor y un sesgo de los genes RP hacia valores más altos de RA que la media (figura 12, D). Ello se debe a la conocida mayor estabilidad los mRNAs de estos genes respecto al promedio del transcriptoma, que causa un aparente exceso de RA para su TR (Grigull, 2004; Pelechano *et al.*, 2010).

1.8- Corrección de datos de GRO y BioGRO por la tasa de *drop off*.

Durante la realización de esta tesis apareció publicado un trabajo del laboratorio de Patrick Cramer (Sun *et al.*, 2012) en el que, mediante la técnica *Comparative Dynamic Transcriptome Analysis*, cDTA, se actualizaba el método de medición de tasas de transcripción maduras (*synthesis rate*, SR) mencionado en el apartado anterior (DTA, Miller *et al.*, 2011). Además de la revisión y el refinado del conjunto de datos de SR, uno de los resultados más interesantes fue el hallazgo de una dependencia inversa entre la SR y la longitud de los genes, de manera que los genes más largos muestran SRs más bajas que los más cortos (adaptado al castellano en figura 13, A). Asumiendo que la tasa de inicios (TI) de transcripción no depende de la longitud del gen, los autores concluyen que la tendencia negativa de la comparación implica que existe un fenómeno de caída, salida o abandono de RNAPs a lo largo del gen (*Drop Off*, en adelante DO), lo cual haría que no todas las RNAPs que empezaran a transcribir llegaran al final del transcrito, a una tasa que asumen constante a lo largo del gen. Teniendo en cuenta por tanto estas dos asunciones fundamentales (TI no dependiente de la longitud del gen, y DO constante a lo largo del transcrito), los autores son capaces de predecir la caída del promedio de SR en función de la longitud, y la cuantifican como un 20% (moléculas/célula/ciclo celular) por Kb. La existencia de una dependencia inversa entre transcripción y longitud de la ORF ya había sido descrita anteriormente (Marín *et al.*, 2003), pero no en el contexto de tasa de transcripción o de síntesis, sino en referencia a la concentración de mRNAs maduros en el citoplasma (RA). Sin embargo, el uso de la RA para el cálculo del DO no es tan apropiada como la SR, ya que la RA es el resultado de la combinación de dos factores: la SR y la tasa de degradación, lo que complicaría más el planteamiento. La SR sin embargo, entendida como la monitorización de la aparición de transcritos recién sintetizados en el citoplasma,

no está influenciada por la DR, con lo que es más fácil relacionarla directamente con la transcripción en el núcleo.

Tanto en el método de GRO radiactivo, en el que la nTR se obtiene a partir de un único valor de señal por gen (1 sonda por gen), como en el de BioGRO, donde se promedia la señal de todas las sondas que cubren su longitud (sondas solapantes de 25 nt), es posible que los valores de nTR de los genes largos estén siendo sobrevalorados, al no tener en cuenta que las señales al principio y al final del gen pueden no ser las mismas debido al DO. Este error es especialmente importante para las nTRs de GRO radiactivo en los genes de longitud mayor de 3kb, debido a que en ellos se usa una sonda correspondiente a la kb final del gen (Alberola *et al.*, 2004). Es por ello que se hace necesaria una corrección de los datos de TR de ambas técnicas, que convierta nTRs *aparentes* (producto del promedio de regiones con distintas densidades de RNAPs activas) en nTRs *reales*.

Como paso previo, y dado que las conclusiones a las que llegan Sun y col. (2012) dependen en gran medida de las dos asunciones mencionadas, usamos los datos de BioGRO para testar una de ellas: TI constante para todos los genes. Como hemos visto, el método BioGRO nos permite tomar valores de nTR de distintos puntos del genoma. Gracias a ello calculamos la TI como el promedio de señal de BioGRO de la zona inicial del gen (+100 a +300 nt desde el TSS), considerando esa región como TI real, ya que en ella las RNAPs han abandonado el promotor, han sido fosforiladas en Ser5, y han pasado a la zona de elongación productiva tras fosforilación en Ser2 (Wade y Struhl, 2008). Los resultados (figura 13, B) muestran que la asunción del trabajo de cDTA no es del todo cierta, ya que se aprecia una ligera bajada de la TI con la longitud del transcrito (6.74% por Kb).

Descontando este dato, la tasa de DO queda en un 13.26%. Con una caída de esa magnitud, se puede extrapolar que la longitud de un transcrito para el que la presencia de RNAPs activas en el momento de terminar la transcripción es el 50% de las iniciales es de 3.8Kb, y extraer, bajo un modelo estocástico, un factor de DO ($\lambda = 0.18$) que aplicar a los datos de nTR de BioGRO para corregirlos, según la ecuación:

$$TR_{real} = \frac{\lambda e^{-\lambda}}{(1 - e^{-\lambda})} TR_{ap}$$

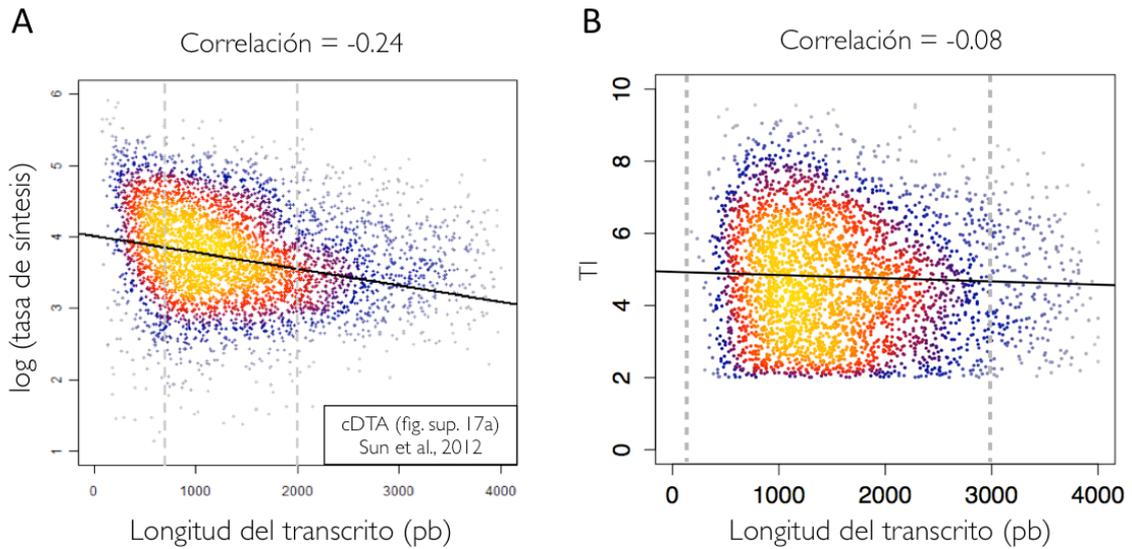


Figura 13. Relación entre la nTR con la longitud de los genes. (A) Comparación entre la SR (aparición de mRNAs en el citoplasma, o tasa de síntesis, medida por el método cDTA) y la longitud de los genes (adaptado de Sun *et al.*, 2012) **(B)** Dependencia de la tasa de inicios (TI) con la longitud de los genes, con datos de BioGRO. Las líneas grises punteadas delimitan en cada caso el rango de tamaños de genes con los que se ha calculado la línea de tendencia: 800 a 2000 en Sun *et al.*, (2012); 300 a 3000 en nuestros datos. TI: promedio de señal de la región de +100 a +300 desde el TSS para todos los genes

Siendo λ el factor estocástico (0.18), l la longitud de cada transcrito, y TR aparente el promedio de señal de GRO/BioGRO para cada transcrito. Aplicando la corrección a los conjuntos de datos de run-on (tanto a los datos de GRO radiactivo (Pelechano *et al.*, 2010) como a los datos de BioGRO), y representándolos en función de la longitud del gen se observa que existe una mayor caída de TR *real* con la longitud del gen (mayor pendiente, figura 14, A y B), así como una mejora de los valores R de correlación. Esta bajada refleja que se ha corregido la sobrevaloración a la que se estaban sometiendo los valores tanto de GRO como de BioGRO, acercándolos a valores de nTR *reales*.

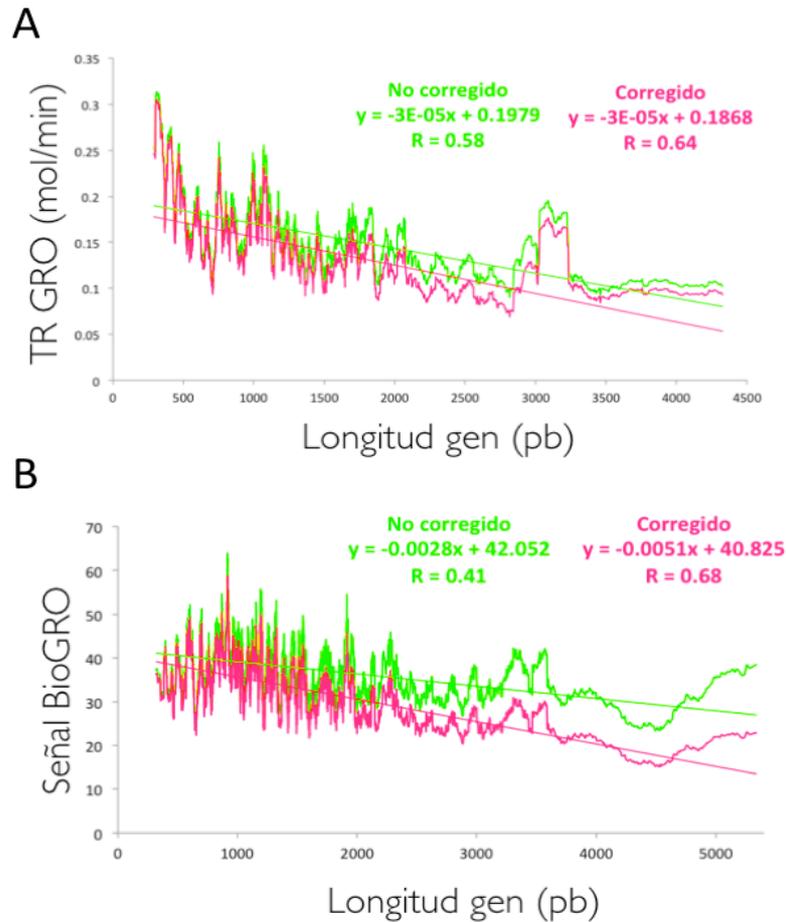


Figura 14. Efecto de la corrección de los datos de TR para tener en cuenta el efecto del Drop Off. (A) Corrección de los datos de nTR medidos por GRO radiactivo (Pelechano *et al.*, 2010). (B) Corrección de los datos de nTR medidos por BioGRO (esta tesis).

Una vez corregidos los datos, quisimos comprobar la validez de la corrección viendo la relación GRO/DTA y BioGRO/DTA en función de la longitud del gen. Las gráficas de ambas comparaciones muestran un cambio hacia la horizontal (no diferencia) de la línea de tendencia tras la corrección (figura 15). Este cambio refleja dos cosas: 1) la línea de tendencia positiva de los datos no corregidos indica que los genes largos están más afectados por el parámetro que diferencia a ambas mediciones (el DO), y 2) los valores de TR nacientes aparentes (run-on) se transforman en TR nacientes reales al tener en cuenta el efecto de DO y corregirlo.

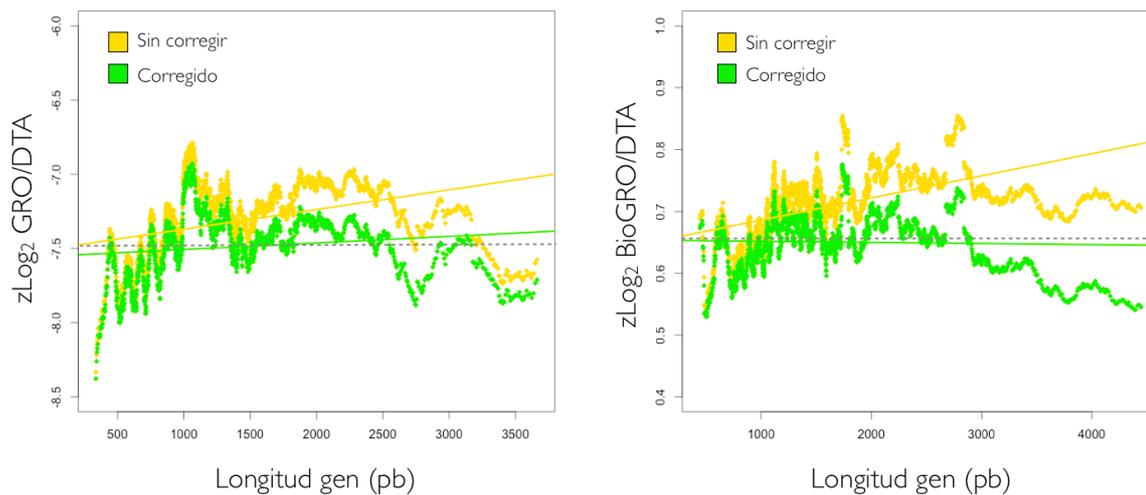


Figura 15. Disminución de las diferencias en las relaciones GRO/DTA y BioGRO/DTA frente a longitud de la ORF tras la corrección de los datos. La línea gris punteada marca la horizontal.

Capítulo 2. Aplicación del método BioGRO a diferentes cepas de levadura. Estudio de cepas mutantes relacionadas con la síntesis y la degradación de RNAs.

Una vez establecidos los fundamentos técnicos del método de BioGRO, el siguiente objetivo que se planteó fue analizar en profundidad los datos para mapear la actividad de las RNAPs activas a lo largo del genoma de la levadura. En esta tesis se han empleado dos maneras generales de representar la presencia de RNAPs activas a lo largo del genoma. La primera consiste en tomar el promedio de las intensidades de todas las regiones de interés y representarlas en un mismo gráfico. Esta estrategia, denominada *análisis del metagen*, ha sido empleada previamente para la representación de datos de transcripción naciente alrededor del TSS y/o del TTS (Core *et al.*, 2008, Nechaev *et al.*, 2010, Churchman y Weissman, 2011, Hah *et al.*, 2011, McKinlay *et al.*, 2011, Kwak *et al.*, 2013). La segunda estrategia es similar a la anterior, aunque difiere en el hecho de que la representación que se realiza no es la de un único perfil promedio de señal, sino que se presentan los perfiles independientes de cada gen, a modo de *mapas de calor*. Hasta donde llega nuestro conocimiento, este tipo de representación únicamente ha sido utilizado previamente en un estudio de transcripción naciente en nematodos (Kruesi *et al.*, 2013). Para ambas representaciones, en esta tesis se tomó como referencia general un conjunto de anotaciones de los TSS y TTS mayoritarios del genoma de *S. cerevisiae* (Xu *et al.*, 2009).

2.1-Gen Modelo: perfil global de densidad de RNAPs activas; caso especial de los genes con intrones.

Mediante la utilización de programación a medida en lenguaje estadístico R (contribuido por Zhenyu Xu, Ishaan Gupta, Vicent Pelechano y Antonio Jordán), se estudió la señal de BioGRO a lo largo del cuerpo del gen y de sus zonas flanqueantes, eligiendo para ello de

manera arbitraria una región que comprende desde 1 kb antes del TSS a 1 kb después del TTS, para todos los genes codificantes de proteínas con anotación (n=5171). De igual manera que para las regiones flanqueantes, la amplitud de región entre el TSS y el TTS se normalizó a una longitud unitaria de 1, que en la escala utilizada en la gráfica es igual a la de 1kb, de manera que lo que se representa es el promedio ajustado a esa región de la densidad real de RNAPs sobre los genes, de manera independiente a su longitud. A este tipo de representación especial del metagen, en la que obtenemos un perfil de toda la región transcrita, la llamamos *Gen Modelo*. En el perfil se aprecia un mínimo de señal de RNAPs activas que coincide con los sitios de inicio y final de transcripción, y un máximo que se mantiene relativamente estable a lo largo del cuerpo del gen, con una ligera acumulación hacia el extremo 3' (Figura 16, A). Estos resultados contrastan con lo que se ha descrito previamente en levadura con una técnica similar (McKinlay *et al.*, 2011), donde se reportó una acumulación de señal de RNAPs activas paradas próximas al promotor. La señal de la transcripción naciente antisentido indica que en la zona transcrita existe menos transcripción de este tipo que en sus flancos. El análisis del *Gen Modelo* nos sirvió para validar a nivel genómico las características particulares de la transcripción naciente descritas para genes individuales en el capítulo 1.

Tal y como se muestra en el anterior capítulo de la tesis (figura 11), el BioGRO es capaz de capturar RNAPs activas sobre las regiones intrónicas de los genes. Es ampliamente conocido que estas regiones son rápidamente procesadas y eliminadas de los mRNAs, en la mayoría de casos de manera co-transcripcional (para una revisión, ver Carrillo Oesterreich *et al.*, 2011), con lo que este hecho evidencia que lo que BioGRO está caracterizando es inequívocamente transcripción naciente. Para demostrarlo a nivel global, llevamos a cabo el análisis del *Gen Modelo* para el subgrupo de los genes con intrones, que representan únicamente alrededor del 5% del total (n=340). El procedimiento computacional fue equivalente al del análisis del *Gen Modelo* de todos los genes. Se modificó para sustituir las coordenadas de TSS y TTS por las de las posiciones de inicio y final del intrón de cada gen, extraídas de la base de datos del genoma de levadura, SGD (<http://www.yeastgenome.org/>). En aquellos casos en los que los genes tuvieran más de un intrón, se analizó solamente la región del primer intrón tras el TSS.

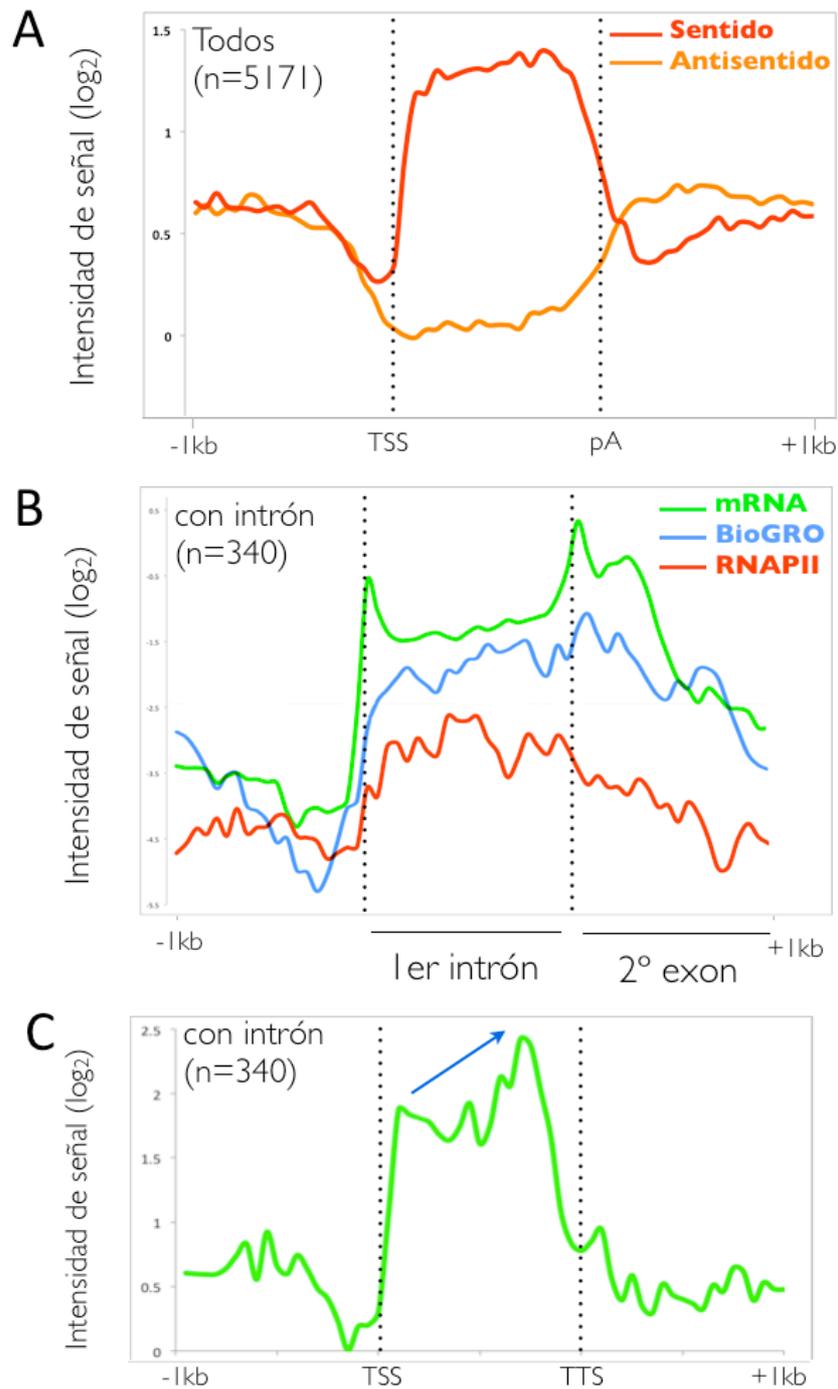


Figura 16. Perfiles promedio de densidad de RNAP activas. (A) Perfil global de densidad de RNAPs elongantes. **(B)** Para los genes con intrones, comparación entre la presencia de transcritos nacientes y maduros en la región del primer intrón y sus zonas flanqueantes. **(C)** Acumulación de RNAPs activas en la región 3' de los genes con intrones. **mRNA:** perfil de mRNA maduro; **BioGRO:** perfil de BioGRO; **RNAPII:** perfil de ChIP de RNAPII (De Miguel y Chávez, no publicado). **Intensidad de señal (\log_2)** es la nTR medida en unidades arbitrarias y representada en escala logarítmica.

A modo comparativo se analizaron de la misma manera datos de un RNA maduro, y de un experimento de CHIP de RNAPII (datos *crudos* cedidos por Sebastián Chávez y Lola de Miguel, Universidad de Sevilla). Los perfiles de los tres conjuntos de datos se obtuvieron por separado y se unieron en una misma gráfica (figura 16, B). Como era de esperar, los resultados muestran una depleción de señal en la zona del intrón, para los datos de RNAs maduros, pero no existe descenso apreciable en las señales de BioGRO o de CHIP de RNAPII, con lo que se confirma que el método de BioGRO está midiendo transcripción naciente, de manera específica, con poca o ninguna contaminación de RNA maduro.

Tal y como se ha descrito recientemente (Carrillo Oesterreich *et al.*, 2010, Alexander *et al.*, 2010), el procesado de intrones en levadura se produce mayoritariamente de manera co-transcripcional. Esta afirmación está basada en el hallazgo por parte de los autores del fenómeno de *Terminal Exon Pausing*, por el cual la RNAPII se pausa alrededor de 250 nt aguas abajo del sitio de corte y empalme durante el tiempo suficiente para que la maquinaria de procesado del intrón termine su actividad. Tras ello, la RNAPII es capaz de reiniciar la transcripción de la parte 3' restante del gen. A la luz de estos datos, nos planteamos testar si nuestros datos eran capaces de reflejar la acumulación de RNAPs hacia 3' de los genes con intrones, para lo cual realizamos el análisis del *Gen Modelo* únicamente con estos genes. Los resultados (figura 16, C), aunque ruidosos por el reducido número de genes promediados, muestran efectivamente una acumulación de RNAPs hacia el final de los genes, no presente en el conjunto de los genes (figura 16, A). Estos resultados son consistentes con la existencia de una parada o frenado de las RNAPs después del intrón, en el segundo exón del gen.

2.2-Análisis del metagen de los extremos 5' y 3' de los genes

Con el objetivo de investigar con más detalle las dinámicas de actividad de RNAPs elongantes en las regiones alrededor del promotor y del terminador de los genes, llevamos a cabo análisis del metagen distinto al descrito en el apartado anterior. En este caso, se toman las señales de los transcritos de interés desde el TSS o el TTS y se representa la densidad de señal en una zona de amplitud variable alrededor de tales

puntos. A diferencia del *Gen Modelo*, este tipo de representación muestra los datos de señal de los transcritos a sus distancias reales.

Del mismo modo que en el apartado anterior, los perfiles de BioGRO generados se analizaron de manera comparativa con los perfiles de un RNA maduro y un ChIP de RNAPII. La densidad promedio de RNAPs elongantes alrededor de la zona de inicio de la transcripción alcanza un valor mínimo en una zona situada a -100 nt del TSS, tras la cual la densidad va aumentando de manera rápida hasta alcanzar un valor máximo poco después del TSS que se mantiene relativamente estable, pero con ligeras fluctuaciones, hasta los 500 pb (figura 17, A). Este perfil difiere claramente del que se obtiene de un RNA maduro, caracterizado, como era de esperar, por un mínimo de señal coincidente con el TSS, y un máximo de señal que se mantiene estable a lo largo de toda la región analizada. En la zona de la terminación también se observa que la densidad de RNAPs activas alcanza un mínimo, al que llega tras un descenso de la densidad ligeramente menos pronunciado que en el TSS, lo que sitúa el mínimo ~180 nt más allá del TTS (figura 17, B). Además de ello, cabe destacar la presencia de un pico de acumulación de RNAPs activas justo antes del TTS, a la que nos referiremos más adelante. El correspondiente perfil de RNA maduro difiere de nuevo en la posición del mínimo de señal, que se sitúa ahora en el TTS, y en la ausencia de fluctuaciones en la región aguas arriba del TTS.

Al analizar con más detalle el perfil promedio de la región 5' de los genes advertimos que las fluctuaciones en la densidad de RNAPs activas tras el TSS seguían una periodicidad de ~165 nt, sugerente de una vinculación con la disposición de los nucleosomas sobre el DNA de las mismas regiones (Lee *et al.*, 2007; Jiang y Pugh, 2009). Se ha descrito que el tránsito de las RNAPs a lo largo de los genes se ve afectado por la presencia de nucleosomas, de tal manera que éstas sufren *backtracking* y paradas en las regiones justo anteriores a las ocupadas por los nucleosomas (Kireeva *et al.*, 2005, Hodges *et al.*, 2009, Jin *et al.*, 2010, Churchmann y Weissman, 2011).

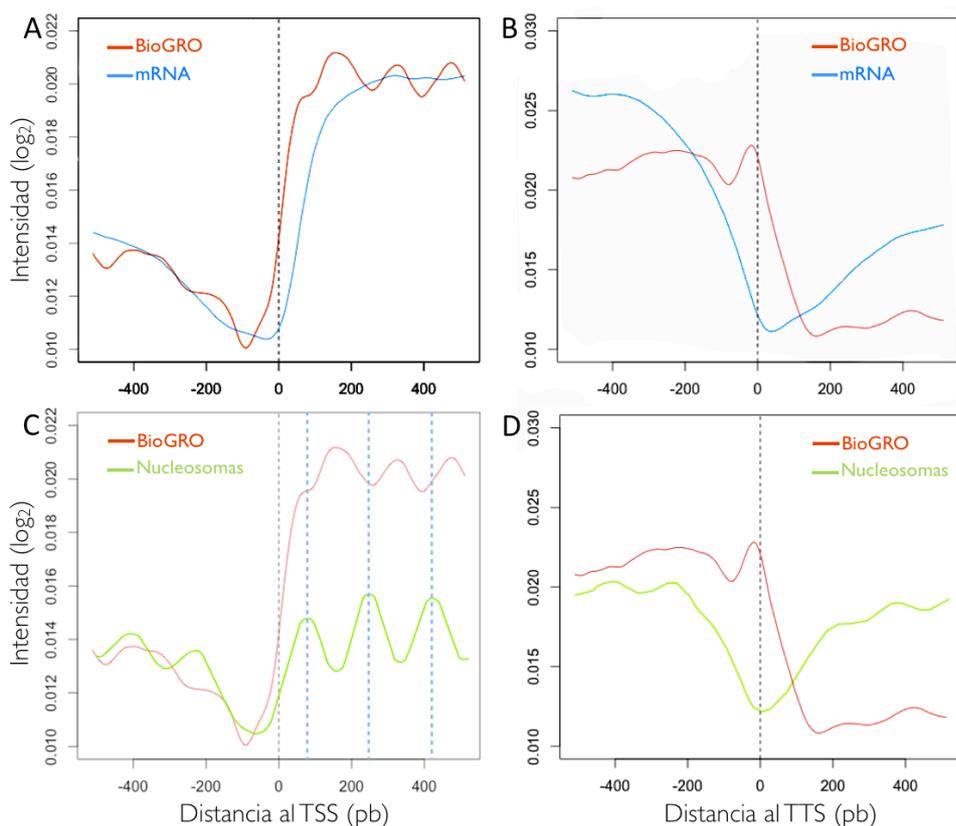


Figura 17. Comparación de perfiles promedio de transcripción naciente con datos de RNA maduro y posicionamiento de nucleosomas. (A) Análisis del metagen desde el 5' para todos los genes. **(B)** Análisis del metagen desde el 3' para todos los genes. **(C)** y **(D)** Superposición del perfil promedio de posicionamiento de nucleosomas para todos los genes (tomado de Weiner *et al.*, 2010). Las líneas punteadas verticales marcan máximos de densidad, correspondientes con las posiciones de los nucleosomas +1, +2 y +3, desde el TSS. **Nucleosomas:** perfil de densidad promedio de nucleosomas (tomado de Weiner *et al.*, 2010). Resto de símbolos como en la figura 16.

De ser así, tales paradas se verían como incrementos de densidad promedio en el perfil obtenido por BioGRO. Para comprobarlo se tomaron perfiles publicados sobre posicionamiento a nivel genómico de nucleosomas, obtenidos mediante secuenciación masiva de DNA (Weiner *et al.*, 2010) y se superpusieron a los de BioGRO a modo de comparación (figura 17, C). Al trazar líneas punteadas que atraviesan el centro de los nucleosomas +1 a +3 se observa la existencia de un patrón de alternancia de máximos y mínimos de los dos perfiles, que sugiere que las RNAPs activas no siguen una dinámica de avance continuo a lo largo de los genes, sino que se acumulan transitoriamente justo

antes de cada nucleosoma. Al superponer el perfil de nucleosomas sobre el perfil de RNAPs en el extremo 3' de los genes no se observa un perfil antinucleosomal evidente (figura 17, D), pero hay que tener en cuenta que los nucleosomas cerca del sitio de terminación no están tan posicionados como en el extremo 5' (Lee *et al.*, 2007, Jiang y Pugh, 2009, Weiner *et al.*, 2010, Brogaard *et al.*, 2012).

2.2.1. Perfil de CHIP de RNAPII en los extremos 5' y 3' de los genes en relación al perfil de RNAPs elongantes.

Para estudiar las posibles similitudes y/o diferencias de los perfiles de RNAP elongantes y RNAP totales sobre los transcritos se realizó un análisis del metagen para los datos de CHIP de RNAPII con anti-Rpb3 (cedidos por De Miguel y Chávez) y se comparó con el de BioGRO (figura 18). A grandes rasgos, el perfil de CHIP obtenido resultó similar al que había sido publicado anteriormente (Mayer *et al.*, 2010), y la superposición de este perfil al de BioGRO y al de nucleosomas (figura 18, A) nos permitió extraer tres conclusiones fundamentales: 1) el perfil de densidad de RNAPII sobre los genes sigue una dinámica similar a la de RNAPs activas: un mínimo de señal ~180 nt antes del TSS, coincidente con la zona libre de nucleosomas (NFR) 5', que sube progresivamente tras el TSS hasta alcanzar un máximo que desciende ligeramente a lo largo de la región estudiada 2) ausencia de patrón nucleosomal de densidad en la región 5', y 3) doble acumulación de RNAPs, antes y después del sitio de pA.

Al analizar con más detalle el perfil de BioGRO alrededor del sitio pA, se aprecia que la primera acumulación (más amplia) está centrada ~ -250 nt del sitio pA, y la segunda (más estrecha), se sitúa justo sobre el sitio pA con un máximo centrado ~ -20 nt. Esta doble acumulación observada puede estar reflejando las dinámicas de formación del extremo 3' y terminación a las que se ven sometidas las RNAPs activas al aproximarse al sitio pA. Es conocido que las RNAPs son alcanzadas al aproximarse al sitio pA por los factores co-transcripcionales de procesamiento (*Co-transcriptional Processing Factors*, CPFs), necesarios para las reacciones de corte endonucleolítico y adición de la cola pA

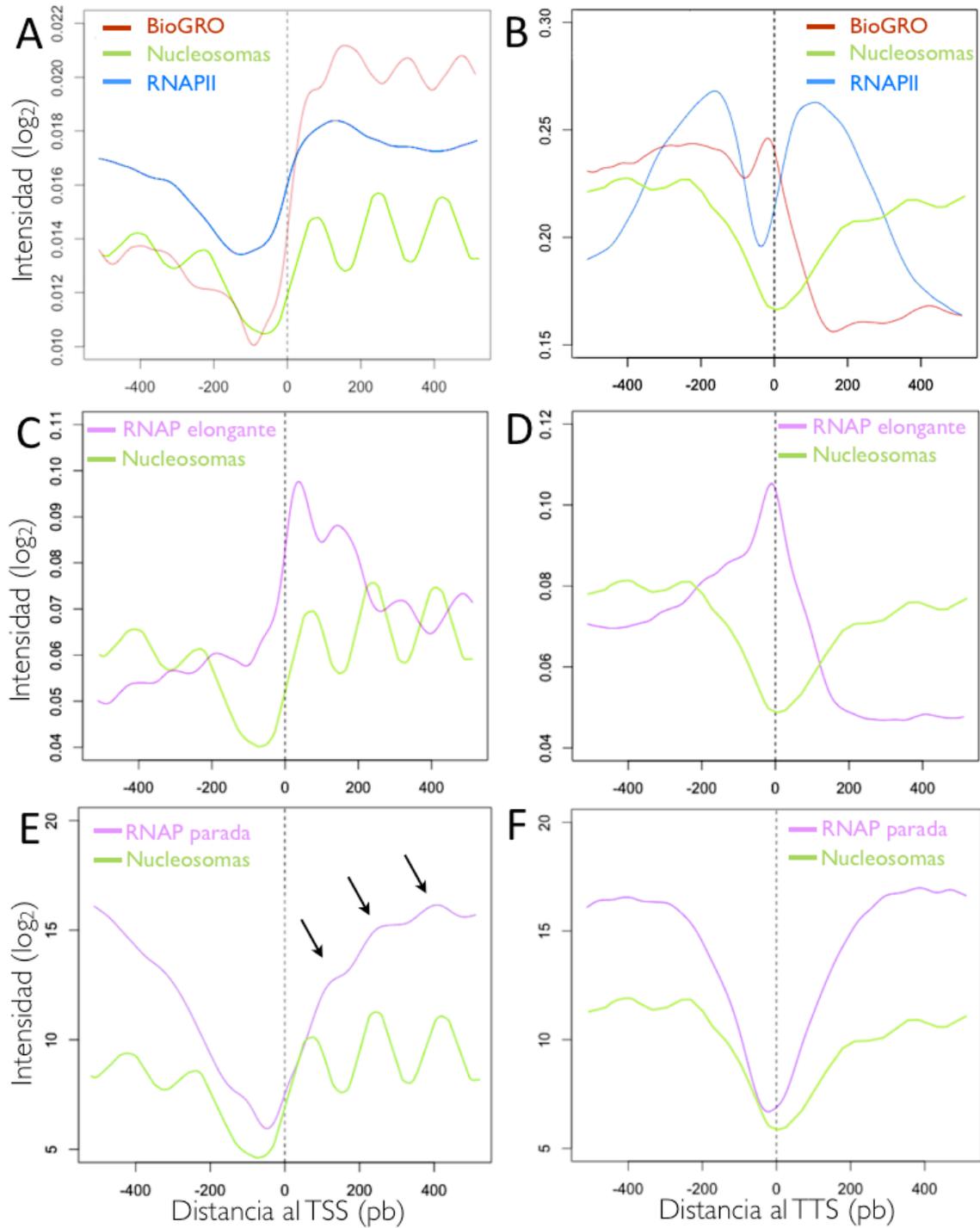


Figura 18. Comparación de perfiles promedio de transcripción naciente con datos de RNAPs totales y posicionamiento de nucleosomas. (A) y (B) Análisis del metagen de los datos de BioGRO, densidad de RNAPs medidas por CHIP con anticuerpo contra la subunidad Rpb3 (De Miguel y Chávez, no publicado) y densidad de nucleosomas (tomado de Weiner *et al.*, 2010), para todos los genes. **(C) y (D)** Perfil resultante de la relación entre los datos de BioGRO y CHIP para todos los genes, alineados por el TSS o el TTS. **(E) y (F)** Perfil de la relación entre los datos de CHIP frente a los de BioGRO. Símbolos como en figuras 16 y 17.

(Ansari y Hampsey, 2005, Mischo y Proudfoot, 2012). Dicha asociación de complejos proteicos puede disminuir la velocidad de la RNAP. Tras ello, algunos factores de elongación se disocian del holoenzima a su paso por el sitio pA, como por ejemplo Spt6, y tras él, las RNAP son alcanzadas por nuevos factores de terminación, como por ejemplo Pcf11 y Rat1 (Mayer *et al.*, 2010). Además de estos factores, la secuencia del DNA de esta región, así como la presencia de nucleosomas también pueden afectar a la velocidad de paso de las RNAPs, reflejándose en el perfil de densidad obtenido por ChIP.

Para destacar mejor el perfil específico de RNAP elongante o de RNAP parada, se dibujaron los perfiles de la relación entre los datos de BioGRO y ChIP. En el perfil de riqueza de RNAPs elongantes (BioGRO/RNAPII, figura 18, B) se aprecia el pico de las RNAPs activas situado antes del nucleosoma +1, que no era visible más que como un hombro anterior al primer pico real en el perfil directo (figura 18, A). En el TTS se aprecia claramente el pico específico de RNAP elongante a ~ -20 pb del pA y se ve que el pico más ancho anterior a -250 pb corresponde a una acumulación general de RNAP, no específicamente las elongantes.

El perfil de enriquecimiento de RNAPs paradas (RNAPII/BioGRO) muestra picos de acumulación coincidentes con la posición de los nucleosomas (señalados con flechas en la figura 18, E), así como mínimos de RNAP parada coincidiendo con los NFRs de 5' y 3'. Lo primero refuerza la idea de que las RNAPs hacen *backtrack* al encontrarse con nucleosomas, y lo segundo sugiere que las RNAPs no hacen *backtrack* durante la poliadenilación, pero sí lo hacen justo después, durante la terminación ($\sim +180$ pb del sitio pA). Por otro lado, la aparente ausencia de RNAPs paradas en las proximidades del promotor está de acuerdo con resultados obtenidos mediante ChIP de diferentes formas elongantes de la RNAP (Mayer *et al.*, 2010), pero contrasta con otros hallazgos en los que sí se encuentra un pico de acumulación de RNAPs paradas en la zona justo después del inicio de la transcripción, obtenidos mediante técnicas similares al BioGRO (Churchman y Weissman, 2011, McKinlay *et al.*, 2011).

2.2.2. Análisis del perfil promedio de RNAPs activas en ciertos subgrupos de genes representativos.

Una vez observada la existencia de un perfil característico de RNAPs activas a lo largo de los genes, quisimos comprobar si todos ellos poseían un perfil similar, o si existían subgrupos concretos con perfiles distintos al resto que pudieran estar enriquecidos en alguna categoría funcional. Para ello decidimos llevar a cabo un análisis de agrupamiento por conglomerados con el método K-medias y una búsqueda de categorías funcionales a partir de los grupos encontrados. Como paso previo se generó un mapa de calor en el que se representaron las intensidades de señal de BioGRO de todos los genes individualmente a lo largo de las mismas regiones 5' y 3' consideradas para el análisis del metagen (figura 19).

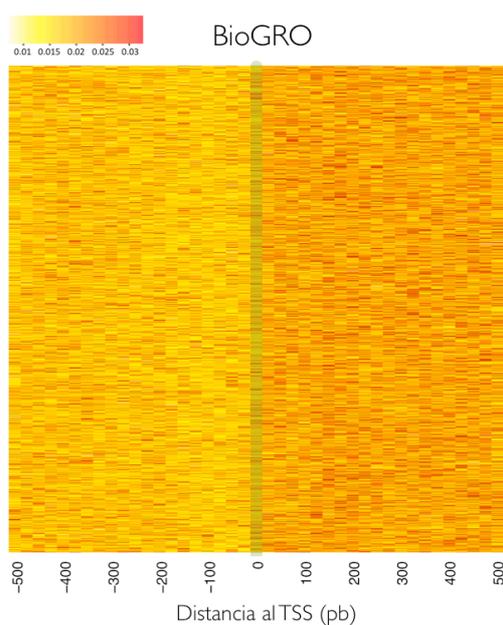


Figura 19. Perfiles individuales de densidad de RNAPs activas alrededor del TSS. Cada línea corresponde a un gen, ordenados por cromosoma y hebra. Código de color en unidades arbitrarias de intensidad (\log_2).

Debido probablemente a la baja intensidad de la señal a nivel de genes individuales, que hace que el nivel de ruido sea alto, las diferencias dentro de los perfiles son poco evidentes, con lo que el método de agrupamiento no consiguió extraer grupos concretos con perfiles similares de manera estadísticamente fiable (datos no mostrados). Se decidió

hacer búsquedas supervisadas de grupos de genes con características funcionales comunes. Se clasificaron los genes según su nivel de expresión en tres categorías de aproximadamente el mismo tamaño, y se dibujaron los perfiles promedio separadamente de cada categoría: BioGRO alto (n=2154), medio (n=2543), o bajo (n=2575) (figura 20, A).

En la figura se observa como los perfiles con fluctuaciones periódicas más nítidos aparecen en los genes con niveles más altos de TR, mientras que el patrón se va perdiendo según va bajando el nivel de expresión y aumenta el ruido técnico. Lo mismo ocurre en la zona alrededor del sitio pA, en la que los genes de alta transcripción son los que contribuyen más al perfil característico observado previamente para todos los genes. Tras ello se analizó el grupo de genes RP. El perfil se caracteriza por un alto valor de señal (~5.7 veces más alto que el promedio de todos los genes, tanto en 5' como en 3'), que va disminuyendo conforme se aleja del TSS, y por la presencia más nítida del mencionado patrón nucleosomal. Alrededor del sitio pA también se aprecia con más claridad la doble acumulación de RNAPs, observada previamente en el conjunto de los genes (figura 20, B). Otro criterio que se utilizó para dividir subgrupos de genes funcionalmente fue su tipo de promotor. De esta forma, se estudiaron los genes cuyo promotor poseyera la secuencia consenso TATA (*Genes TATA*, n=918), o no la poseyera (*Genes TATA-less*, n=2289) (figura 20, C). El perfil de los genes sin caja TATA en su promotor no difiere sustancialmente del perfil global, tanto en el TSS como en el pA. No así los genes con caja TATA, que presentan un mayor nivel de densidad de RNAPs (significativamente mayor al del resto de genes, tabla 3) y una subida de la señal desde el TSS hacia dentro del gen y bajada progresiva hacia el sitio de pA

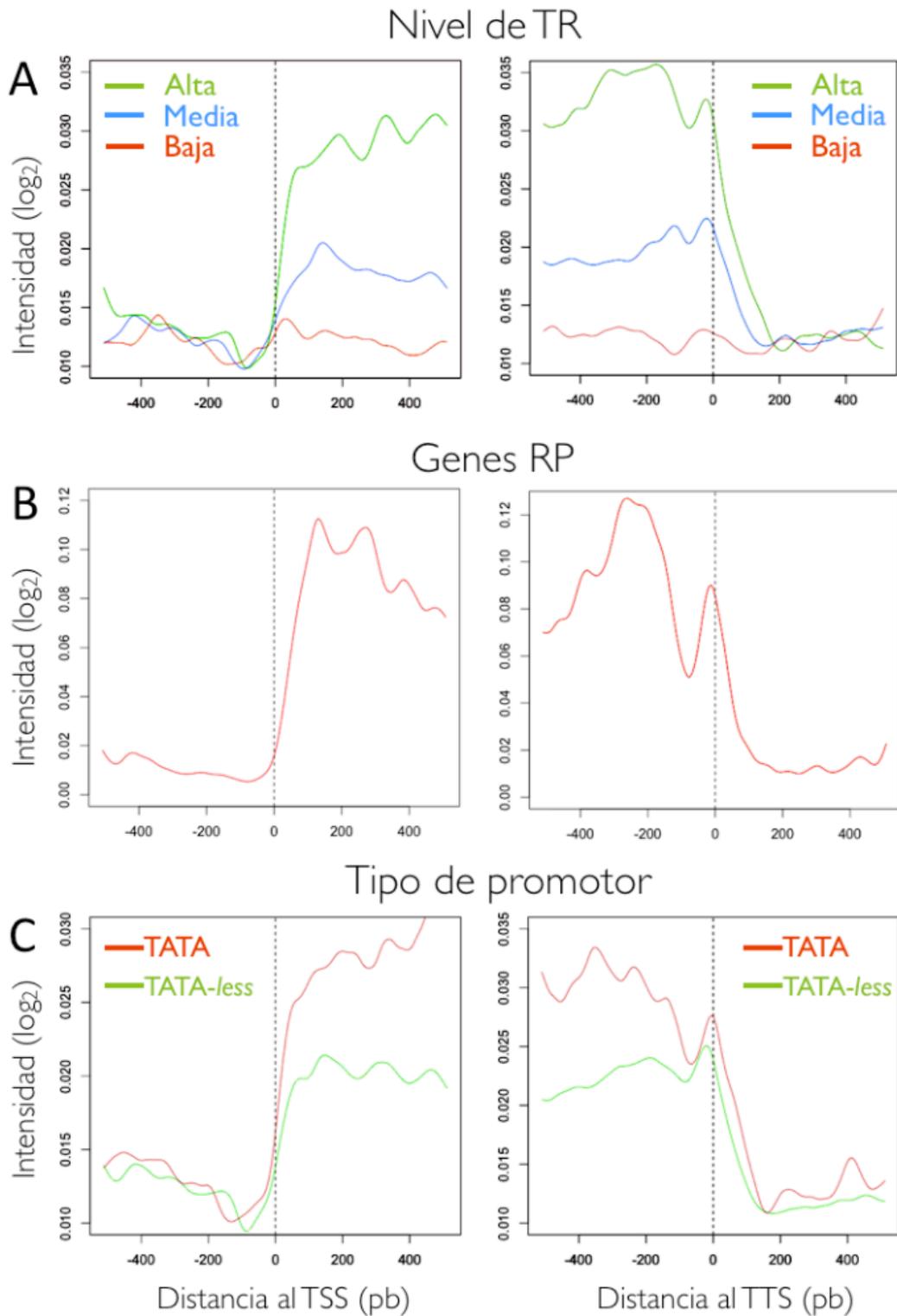


Figura 20. Perfiles promedio de densidad de RNAP activas para distintos grupos de genes. (A) Densidad de RNAPs activas según el nivel de TR de los genes. **(B)** Perfil promedio de transcripción naciente en la región de los genes RP ($n=137$). **(C)** Densidad de RNAPs elongantes en genes con ($n=918$), o sin caja TATA ($n=2289$) en el promotor.

	RA		GRO		BioGRO	
Tipo de promotor	TATA	TATA-less	TATA	TATA-less	TATA	TATA-less
Promedio	0.29	0.1	0.13	-0.009	-7.72	-7.85
Desviación	1.01	0.84	1.26	0.91	1.58	1.16
p-valor	0.0001*		0.0035*		0.03*	

* estadísticamente significativo 95%

Tabla 3. Niveles de expresión según el tipo de promotor. Los valores son en escalas diferentes y en unidades arbitrarias. Procedencia de los datos: BioGRO (esta tesis); GRO (Pelechano *et al.*, 2010); RA (datos de concentración de mRNAs celulares, Wang *et al.*, 2002).

De hecho, los genes con caja TATA presentan mayor nivel de transcripción, tanto en datos de BioGRO como de GRO *clásico* (Pelechano *et al.*, 2010) y de RA (Wang *et al.*, 2002). Cabe destacar también la aparición de fluctuaciones con patrón de espaciado regular en la región caracterizada como acumulación amplia de RNAPs a ~ -250 pb del pA en el perfil de todos los genes (figura 17, B), coincidente con la zona de la primera fase de la transición 3'. A pesar de ello, cabe tener en cuenta que el número relativamente más bajo de los genes TATA contribuye a que presenten un perfil más ruidoso, en el que es más difícil sacar conclusiones.

Estos resultados a escala genómica, junto con hallazgos previos sobre acumulaciones de RNAPII en los extremos 5' y 3' de determinados genes o grupos de genes (Kim y Rando, 2010, Rodríguez-Gil *et al.*, 2010), indican que la densidad de RNAPII dentro de las ORFs presenta un carácter marcadamente variable, que probablemente sea el reflejo de las necesidades regulatorias particulares de los diferentes grupos funcionalmente relacionados.

2.2.3. Análisis de la presencia de RNAPs activas más allá de los límites canónicos de los transcritos.

Tal y como se ha mostrado en el apartado 2.2.2, los datos de BioGRO desvelan la existencia de RNAPs elongantes más allá de los límites estrictos de inicio y final de transcripción (figura 17, A y B). Para analizar mejor este fenómeno y tratar de evaluar su extensión, decidimos hacer un estudio comparativo de los datos de BioGRO y de mRNA. La muestra de mRNA nos sirvió de control negativo para la presencia de señal más allá de los sitios de inicio y poliadenilación. Para ello dibujamos un mapa de calor alineando los transcritos por su TSS y TTS y ordenándolos por su tamaño, de menor a mayor. Una vez obtenidos los mapas de calor de las dos muestras, se superpusieron con un software de análisis de imagen de manera que la intensidad de color de la imagen resultante fuera la relación entre BioGRO y mRNA (figura 21).

Por una parte, el mapa de calor de la muestra de BioGRO, a pesar de la baja señal general de hibridación, muestra una clara separación entre zonas génicas e intergénicas. Por otra parte, en la muestra de mRNA, en la que la intensidad de la señal global es considerablemente mayor, se aprecian con claridad tanto las divisiones entre regiones transcritas y no transcritas, así como las zonas de señal por debajo del fondo, que coinciden con las zonas NFR características de las regiones del promotor y del terminador de los genes. Para realzar la diferencia entre ambos mapas dividimos la señal de BioGRO por la de mRNA. En la gráfica de la derecha se pueden observar zonas más claras en los extremos de los transcritos que corresponden a las regiones con mayor señal relativa de BioGRO respecto al mRNA, que se extiende a lo largo de ~60 pb antes del TSS y ~125 pb después del TTS.

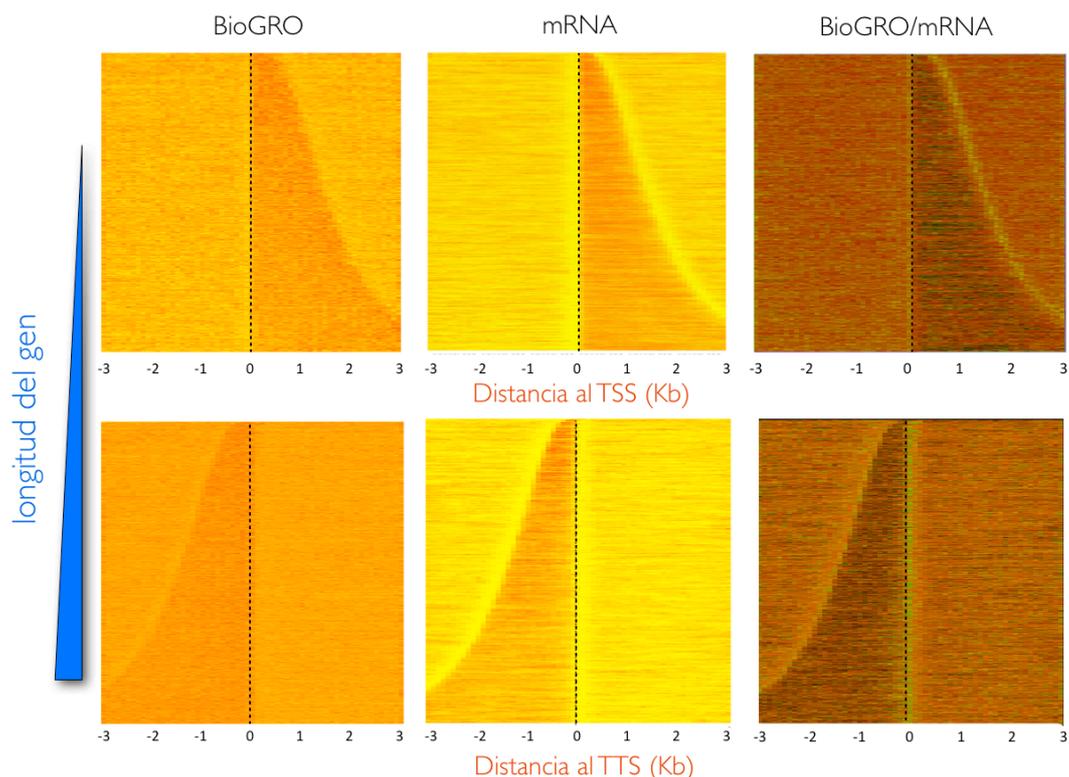


Figura 21. Comparación entre datos de transcripción naciente y madura alrededor del TSS y del TTS a nivel de genes individuales, ordenados por tamaño. Los tres paneles superiores muestran los datos alineando por el TSS y los inferiores por el TTS. Las zonas más claras de los dos paneles de la derecha reflejan la acumulación de señal de RNA naciente frente a maduro.

La presencia de RNAPs activas anteriores al TSS puede explicarse por la existencia de eventos de transcripción naciente en el mismo sentido de transcripción que hacen uso de TSS alternativos (Pelechano *et al.*, 2013). En el caso de la señal más allá del TTS, puede influir tanto la existencia de sitios múltiples de pA (Wilkening *et al.*, 2013; Moqtaderi *et al.*, 2013), como el propio proceso de terminación de la transcripción, que incluye una etapa en la que las RNAPs elongan hasta una distancia variable tras el sitio pA a la espera de ser alcanzadas por el complejo de terminación, en el que está implicada la proteína Rat1 (Richard y Manley, 2009).

Al inspeccionar con más detalle la figura de la relación BioGRO/mRNA alineado por el sitio pA, nos dimos cuenta de que la distancia a la que llega la acumulación de señal más allá del sitio pA parece comparativamente mayor para los genes más cortos. Para comprobarlo decidimos subdividir el conjunto de datos en subconjuntos de genes

según su longitud y dibujar sus perfiles del metagen 3' (figura 22). El panel A de la figura muestra el perfil global de todos los genes, tanto como mapa de calor (izquierda), como de metagen (derecha). Al dividir los genes arbitrariamente en cortos, medios y largos (según se explica en el pie de figura), no parece haber diferencia en la densidad de RNAPs más allá del sitio pA (figura 22, B). Sin embargo, si seleccionamos únicamente los genes más cortos (0-0.5 kb) vemos como el perfil sí se extiende más allá del sitio pA que el conjunto de genes.

A la luz de nuevas caracterizaciones de las isoformas del transcriptoma de *S. cerevisiae* (Pelechano *et al.*, 2013), este resultado podría estar reflejando el hecho de que las RNAPs estuvieran transcribiendo más de un gen en una sola ronda de transcripción. Se ha descrito que esta transcripción podría ser solapante con el gen que se encuentra aguas abajo de manera parcial (terminando 100-200 pb desde el inicio del gen en tándem) o total (produciendo un mensajero bicistrónico). Encontramos que, de los 131 genes que componen el grupo de interés, 70 (53.43%) poseían un gen aguas abajo en el mismo sentido de transcripción. Debido a que este ligero enriquecimiento resultó ser estadísticamente significativo, decidimos visualizar si había diferencias en el perfil del metagen 3' de los distintos grupos de genes, tándem y no tándem, pero no se encontró ninguna diferencia en cuanto a la extensión de la señal de BioGRO más allá del sitio de pA (datos no mostrados).

En otros organismos eucariotas se han hallado evidencias de acumulaciones de RNAPs más allá del sitio pA. Estas acumulaciones se relacionan con parámetros como el nivel de transcripción y el tamaño del gen; siendo los genes más cortos y más expresados los que muestran una mayor acumulación (Rita Grosso *et al.*, 2012). En levadura, los genes cortos también muestran un nivel general mayor de TR, con lo que nuestros resultados apoyan la hipótesis de que la presencia de RNAPs más allá del sitio de pA obedece al nivel de transcripción, siendo mayor en los genes más cortos.

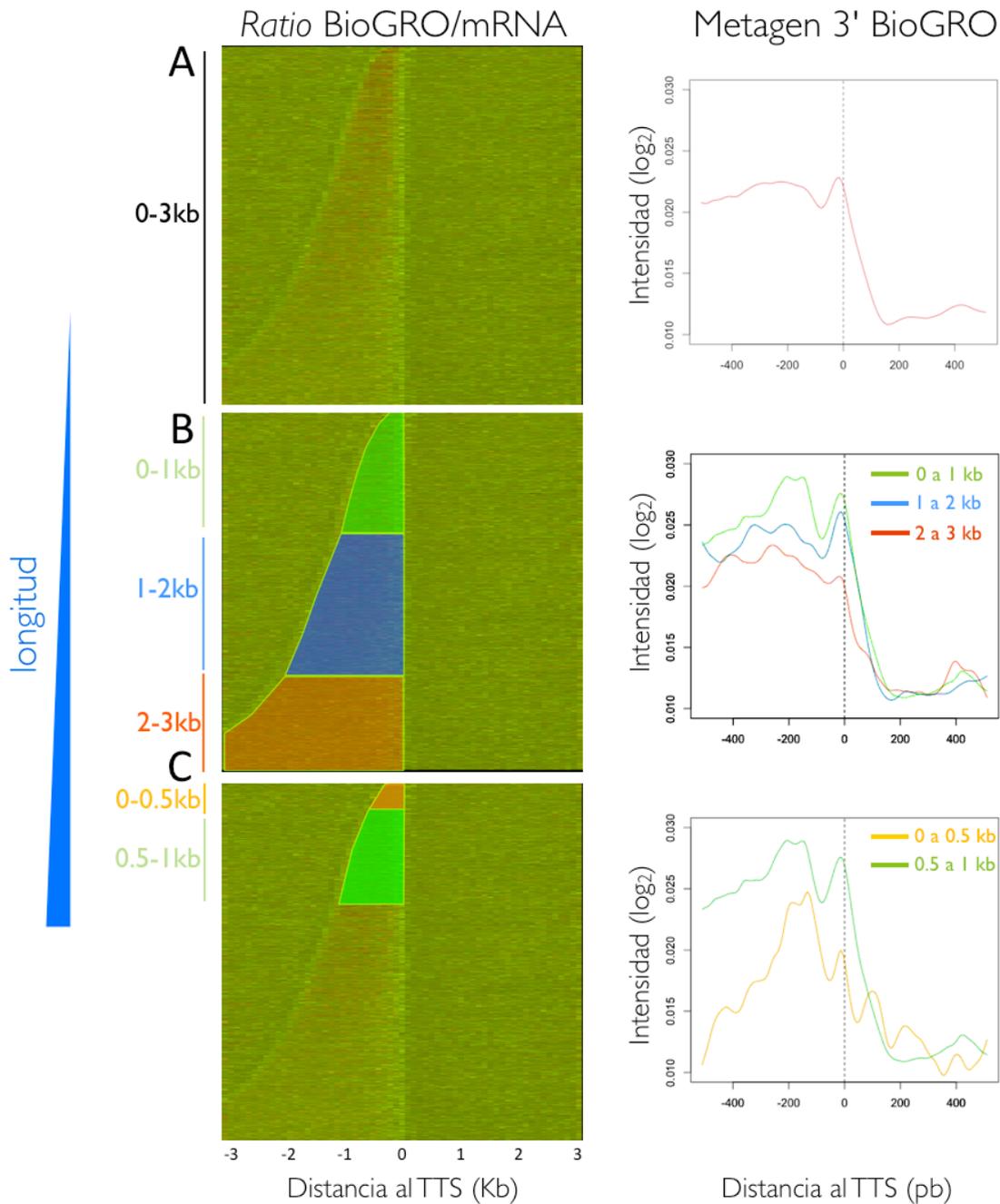


Figura 22. Análisis de la presencia diferencial de RNAPs elongantes más allá del TTS en relación a su tamaño. (A, izquierda) Relación de intensidades de señal de BioGRO y mRNA de todos los genes ($n=4577$), ordenados por tamaño, alrededor del TTS. **(A, derecha)** Análisis del metagen 3' de los datos de BioGRO para esos genes. **(B, izquierda)** Subgrupos realizados según el tamaño del gen: genes cortos (0-1kb, $n=1214$), genes medios (1-2kb, $n=2342$), y genes largos (2-3kb, $n=1021$). **(B, derecha)** Perfil promedio centrado en TTS para cada subgrupo. **(C)** Los genes cortos se subdividieron en genes muy cortos (0-0.5kb, $n=131$) y resto (0.5-1kb, $n=1083$). La figura de la derecha muestra el perfil de ambos subgrupos desde el TTS.

Por otra parte, existe la posibilidad de que la presencia de RNAPs activas en esas regiones esté relacionada con el mecanismo del *gene looping*, por el cual las RNAPs pasan del final de los genes de nuevo al promotor para comenzar un nuevo ciclo de transcripción (Ansari y Hampsey, 2005), siendo necesaria la transcripción de cierta extensión del extremo 3' para que se produjera la unión eficiente entre las regiones de DNA terminador-promotor (Kuehner *et al.*, 2011).

2.3-Estudio del nascentoma del mutante *rpb4Δ*.

Una vez analizados con detalle los datos de BioGRO de una cepa salvaje de levadura, decidimos aplicar el método al estudio del nascentoma de una cepa mutante en un factor relacionado directamente con la transcripción: la cepa *rpb4Δ*. Esta cepa es defectiva en una de las dos subunidades pequeñas no esenciales de la RNAPII. La subunidad Rpb4 se asocia *in vivo* con la otra subunidad pequeña, Rpb7, y juntas componen un submódulo funcional al que se le han atribuido papeles diversos en la célula: por una parte como activador de la transcripción (Woychik *et al.*, 1989, Rodríguez-Gil *et al.*, 2010), y por otra parte como estimulador de la degradación de mRNAs. Es por ello que el submódulo Rpb4/7 es considerado un coordinador de mRNA (Harel-Sharvit *et al.*, 2010, Shalem *et al.*, 2011).

Para obtener los datos de TR naciente y los perfiles de densidad de RNAPs a lo largo de los genes se procedió de la misma manera que con la cepa salvaje. Para la comparación de los datos de TR del mutante con los de la cepa salvaje se eligieron solamente aquellos genes cuyos valores de intensidad de señal estuvieran por encima del umbral de ruido técnico de los microchips. El número de genes que se comparó fue por tanto de 810 (figura 23, A). Resultados previos de nuestro grupo basados en la aplicación de la técnica de GRO al mutante *rpb4Δ* muestran que, al hacer la relación entre las TRs de *rpb4Δ*/WT y representarla frente a la longitud de los genes, existe un defecto de elongación mayor cuanto más largo es un gen (Daniel Medina, comunicación personal). Para comprobar si el mismo fenotipo se observaba mediante la técnica de BioGRO, llevamos a cabo la misma comparación. Los resultados muestran, efectivamente, la existencia de tal defecto,

señalado mediante una línea negra continua que muestra la diferencia entre la tendencia general y la local para los genes cortos (figura 23, B).

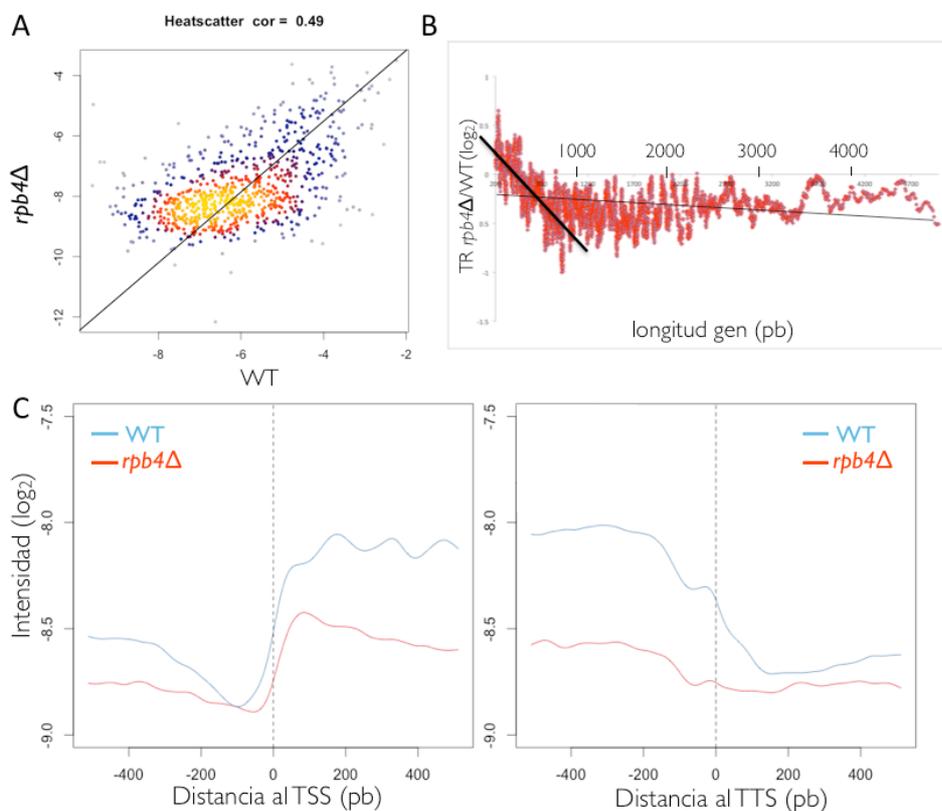


Figura 23. Aplicación del método BioGRO al estudio de la transcripción nascente en el mutante *rpb4Δ*. (A) Correlación entre las TRs de la cepa silvestre y la cepa mutante. (B) Relación de los valores de nTR con respecto al WT, en relación a la longitud de los genes. (C) Comparación de los perfiles de RNAPs activas alrededor de los sitios de inicio y final de la transcripción.

El análisis del metagen del extremo 5' también muestra un perfil de *rpb4Δ* que difiere sensiblemente del de la cepa salvaje, observándose una menor subida de señal tras el TSS y un perfil de caída progresivo a partir de ~180 nt (figura 23, C). Esta caída de densidad puede deberse a dos posibles causas: 1) que las RNAPs empiecen a transcribir pero no todas pasen de la fase de iniciación a la de elongación productiva por la ausencia de la subunidad Rpb4, 2) que no se trate en realidad de una caída de RNAPs del DNA, sino de un aumento de velocidad de las mismas conforme se alejan de las zonas más próximas al

promotor. El perfil en el extremo 3' sigue un patrón similar al del WT, siendo la única diferencia el menor nivel global de señal obtenido para el mutante.

2.4- Estudio del nascentoma del mutante *xrn1Δ*.

Evidencias recientes demuestran la participación de los factores citoplasmáticos de degradación de los mRNAs en la transcripción. Esta participación puede ser indirecta, mediante el control de la degradación de transcritos antisentido que modulan la expresión de ciertos grupos de genes (van Dijk *et al.*, 2011); o bien directa, mediante procesos que implican la entrada de los factores de degradación al núcleo, su unión a la cromatina alrededor del TSS y la estimulación de las fases de inicio y elongación de la transcripción (Haimovich *et al.*, 2013).

Para investigar más a fondo los posibles papeles de factores de degradación sobre la densidad de RNAP en los genes, decidimos caracterizar mediante BioGRO el mutante *xrn1Δ*, deficiente en el principal factor de degradación de los RNAs celulares (Garneau *et al.*, 2007; Parker, 2012; Pérez-Ortín *et al.*, 2013). El perfil global de densidad de RNAPs que se obtuvo por BioGRO con esta cepa fue muy débil y ruidoso, con lo que se pueden extraer pocas conclusiones claras al respecto. A pesar de ello, y como resultado preliminar a falta de aumentar la sensibilidad del protocolo para caracterizar en profundidad cepas como *xrn1Δ*, tan afectadas en transcripción, podemos decir que los perfiles muestran una bajada del nivel de RNAPII respecto del WT sobre el conjunto de los genes, tanto en el extremos 5' como 3' (figura 24, A). Además, el perfil promedio de densidad sobre las regiones antisentido de cada gen aparece aumentado respecto al WT (figura 24, B), específicamente en la regiones de la ORF, lo cual apunta a que la transcripción antisentido desde el NFR 3' está aumentada de manera global. Además, cuando miramos el nivel de ocupación de RNAPs activas sobre los 1658 transcritos XUTs (*Xrn1-sensitive Unstable Transcripts*; caracterizados por van Dijk *et al.*, 2011), observamos un aparente aumento respecto a la cepa WT (figura 24, C).

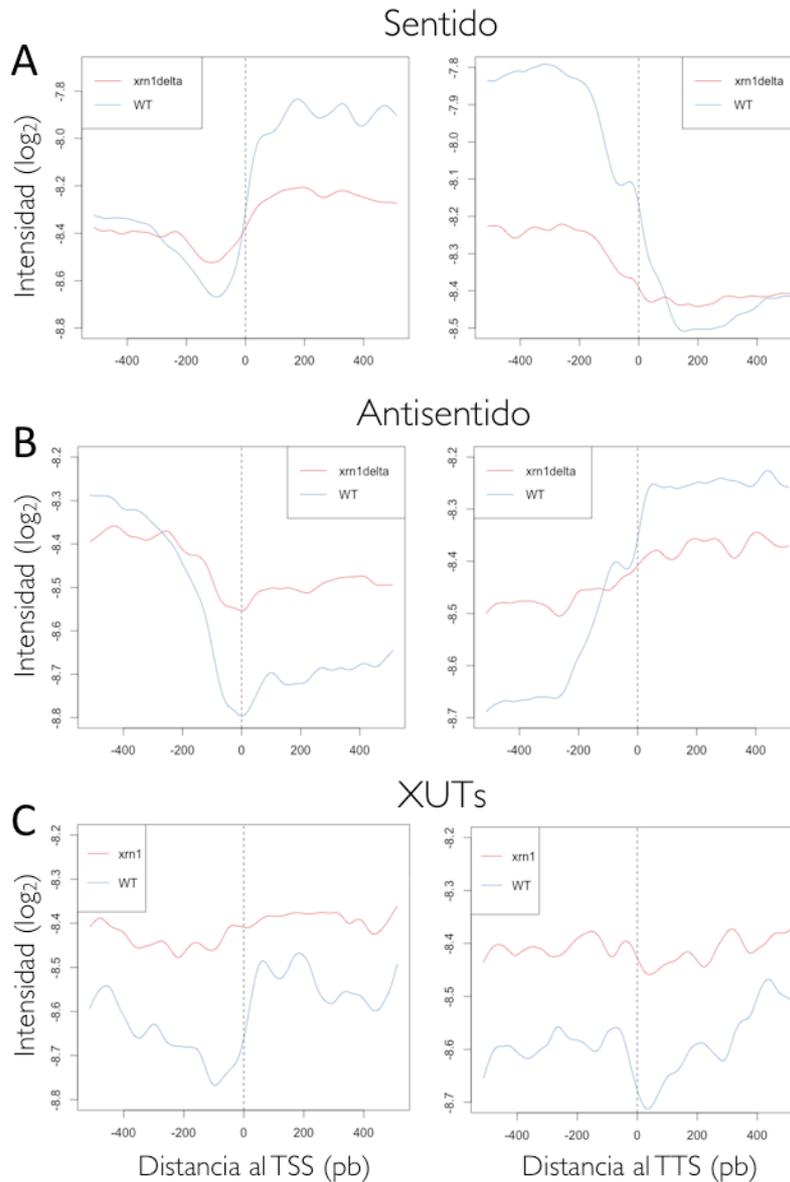


Figura 24. Efecto de la ausencia de Xrn1p en la densidad de RNAPs elongantes a escala genómica. (A) Comparación de los perfiles promedio de transcripción nascente. **(B)** La misma comparación, pero tomando los datos en antisentido de cada región. **(C)** Niveles de expresión de los transcritos XUT en las dos cepas.

En conjunto, estos resultados indican que la ausencia de Xrn1 compromete la transcripción afectando la ocupación de RNAPII sobre los genes, a escala genómica. Este efecto es uniforme a lo largo de la región transcrita, y parece conllevar un aumento en los niveles de transcripción antisentido. Tomados en conjunto, y teniendo en cuenta su carácter preliminar, estos resultados sugieren que la exonucleasa Xrn1p ejerce un control

sobre los transcritos no codificantes XUTs, tanto a nivel de su degradación en el citoplasma, como directamente sobre su transcripción en el núcleo.

Capítulo 3. Relación entre RNAPs activas y cromatina

Tal y como se ha comentado en el capítulo anterior, la disposición de las RNAPs activas mapeadas mediante BioGRO a lo largo de los genes no sigue una distribución uniforme, sino que presenta heterogeneidades dependientes de la zona del gen. En el caso concreto de la zona 5' de los genes, las RNAPs elongantes presentan fluctuaciones periódicas de densidad, coincidentes en amplitud, pero alternadas espacialmente, con la disposición de los nucleosomas sobre el DNA. Tal disposición hace pensar en una posible relación causal entre la posición que ocupan los nucleosomas y la densidad de RNAPs activas, la cual ha sido previamente demostrada tanto *in vitro* (Fish y Kane, 2002, Hodges *et al.*, 2009), como *in vivo* (Churchman y Weismann, 2011). La demostración *in vivo* de la relación entre actividad de RNAPs/presencia de nucleosomas se basa en la aplicación de una técnica, NET-seq, similar al BioGRO. La técnica de NET-seq se fundamenta en el aislamiento, secuenciación y mapeo sobre el genoma de levadura de los transcritos nacientes procedentes de la IP del complejo ternario DNA-RNAP-RNA naciente. Tal aproximación refleja por lo tanto la presencia del complejo de elongación de la transcripción, independientemente de su actividad elongante. Mediante el uso de la técnica de BioGRO, que mide exclusivamente las RNAPs activas, se puede tener una visión alternativa y complementaria del impacto que tienen los nucleosomas en la transcripción. Para ello, y como objetivo de este capítulo de la tesis, se llevó a cabo un estudio detallado de la relación entre nucleosomas y RNAPs activas, que comprende tanto el reanálisis bioinformático de datos publicados de posicionamiento de nucleosomas, como el empleo de la técnica de BioGRO en cepas de levadura carentes de factores relacionados directamente con el proceso de la transcripción en su contexto cromatínico.

3.1-Análisis de datos de nucleosomas.

Los datos crudos de posicionamiento de nucleosomas, obtenidos mediante la hibridación de DNA mononucleosomal en microchips de embaldosado de Affymetrix (Whitehouse *et*

al., 2007) se descargaron de la página web del laboratorio del Profesor Tsukiyama (http://research.fhcr.org/tsukiyama/en/microarray_data.html), y se reanalizaron con nuestros algoritmos. Una vez procesados, los datos se utilizaron para generar una representación del tipo mapa de calor, de igual manera a como se describe anteriormente aplicada a los datos de BioGRO (capítulo 2, figura 19). En este caso lo que se representó fue la densidad de nucleosomas en una región fija de 1000 pb alrededor del TSS para todos los genes de levadura. En el gráfico resultante se pueden apreciar bandas más oscuras, correspondientes a regiones cubiertas por nucleosomas, y regiones más claras, correspondientes a regiones desprovistas de nucleosomas (figura 25, A).

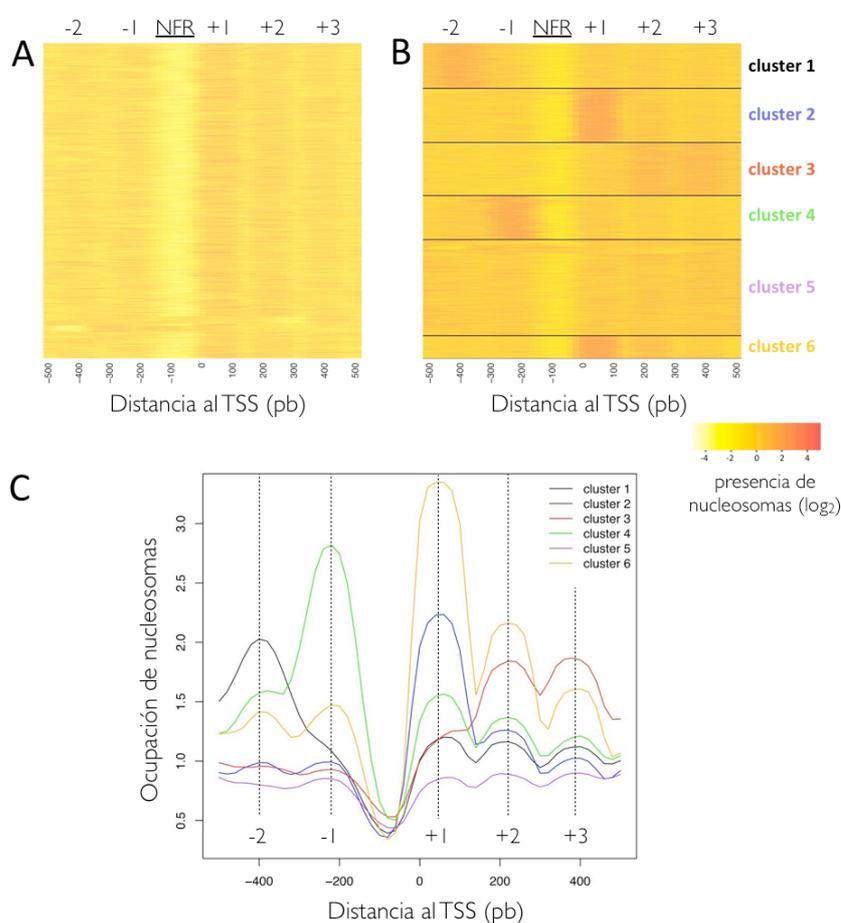


Figura 25. Reanálisis de datos de posicionamiento global de nucleosomas. (A) mapa de calor de posicionamiento de nucleosomas para cada gen. **(B)** mapa de calor del agrupamiento por similitud de perfil mediante k-medias. **(C)** perfiles promedio de posicionamiento de nucleosomas alineados por el TSS para los genes de cada grupo encontrado en (B). En el interior del gráfico se marca la nomenclatura y el centro de cada nucleosoma.

Cabe destacar una de estas zonas, más amplia y situada alrededor del TSS, correspondiente a la zona libre de nucleosomas (NFR) ya mencionada, característica de la región promotora de los genes.

Para extraer conjuntos de genes con patrones parecidos de nucleosomas se agruparon los datos en 6 grupos mediante el método k-medias (figura 25, B). De cada uno de los grupos establecidos se extrajo la lista de genes que lo componía y con ella se generó una representación tipo metagen para ver el perfil de densidad de nucleosomas promedio (figura 25, C). Los resultados muestran que existen diferentes grupos de genes con patrones claramente diferenciados de nucleosomas en sus regiones 5': algunos caracterizados por tener el nucleosoma +1 muy bien posicionado respecto al resto, otros poseen el nucleosoma -1 más posicionado que el resto, y otros carecen de un nucleosoma +1 o -1 posicionado pero destacan por el posicionamiento relativo de los nucleosomas más distales al TSS. Tenemos por tanto:

- Grupos 2 y 6: nucleosoma +1 bien posicionado.
- Grupo 4: nucleosoma -1 bien posicionado.
- Grupos 1 y 3: nucleosomas distales al TSS bien posicionados.
- Grupo 5: patrón nucleosomal débil, correspondiente al grupo mayoritario de genes. Contiene un subgrupo de genes cuyo NFR parece estar ocupado por un nucleosoma.

Para ver si existía alguna relación entre los diferentes patrones de posicionamiento de nucleosomas encontrados y los de RNAPs activas se dibujó el perfil promedio de BioGRO para cada uno de los grupos, y los perfiles se superpusieron en un mismo gráfico. Los resultados muestran como un determinado patrón de posicionamiento nucleosomal va acompañado de un perfil de RNAPs activas determinado, siendo los más claros los obtenidos al comparar los perfiles de BioGRO de los grupos de genes con marcadas diferencias de posicionamiento del nucleosoma +1 (figura 26). Así, los genes del clúster 1 se caracterizaron por poseer una elevada señal de BioGRO general asociada a un perfil nucleosomal débil, mientras que los del clúster 2 poseen una señal más baja

comparativamente en la región del nucleosoma +1, asociada a un perfil nucleosomal en el que éste es el nucleosoma mejor posicionado.

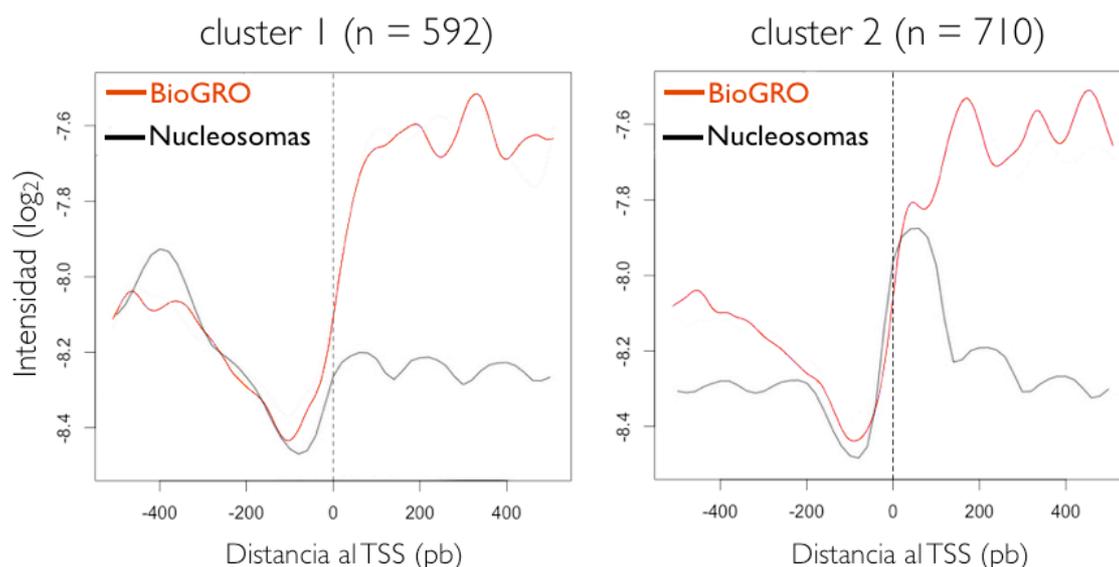


Figura 26. Relación entre la estructura de la cromatina y la actividad de la RNAPII. Los gráficos muestran superpuestos los perfiles de los genes de los grupos 1 y 2 derivados del análisis de agrupamiento por k-medias (figura 25, B y C).

Los resultados están de acuerdo con la hipótesis de que los nucleosomas dificultan de manera notable el paso de las RNAPs a lo largo de las regiones a transcribir.

3.2-Análisis de la cepa mutante *isw2Δ*.

Los resultados hasta este punto ponen de manifiesto que la estructura cromatínica de los genes tiene un impacto directo sobre la transcripción de la RNAPII. En condiciones normales, es capaz de determinar la densidad de RNAPs a lo largo de regiones concretas de los genes (figura 26).

Para tener una confirmación directa de la relación entre nucleosomas y RNAPs decidimos caracterizar mediante BioGRO el nascentoma del mutante *isw2Δ*, para el

que se había descrito en la literatura que presentaba un patrón alterado de nucleosomas. La proteína Isw2p forma parte del complejo ISWI de remodeladores de la cromatina con actividad ATPasa. Se ha descrito que Isw2p se une a las regiones que flanquean los NFRs de ciertos genes y reposiciona los nucleosomas acercándolos a los NFR (Whitehouse *et al.*, 2007, Yen *et al.*, 2012). En ausencia del factor remodelador, los nucleosomas adoptan una posición ~10 pb en promedio más alejada de los NFR (Yen *et al.*, 2012), marcada mayoritariamente por la secuencia del DNA subyacente.

Como resultado general cabe destacar la aparición de un defecto de elongación progresivo, que a diferencia de lo que ocurre con otros mutantes caracterizados en esta tesis (*rpb4Δ* y *dst1Δ*), aparece como una disminución progresiva de la TR con la longitud del gen (figura 27, B). De esta manera se concluye que el defecto de elongación es más acusado conforme aumenta la longitud de la zona a transcribir. No se aprecian diferencias sustanciales de perfil de metagen 5' y 3' respecto al WT (figura 27, C), salvo por una ligera bajada de intensidad de señal general. Pensamos que estos resultados entran dentro de lo esperable, ya que el factor Isw2 se une aproximadamente a un 12% de los genes (Yen *et al.*, 2012), con lo que el posible cambio de posicionamiento puede quedar enmascarado al promediar su perfil con el del resto de genes de la levadura, cuyas posiciones no han cambiado.

Para refinar la búsqueda del fenotipo se trabajó con dos listas de genes, derivadas de los estudios previos de caracterización del mutante *isw2Δ* publicados en la literatura. La primera lista está compuesta por genes cuyo patrón de posicionamiento de nucleosomas se ve alterado en el mutante *isw2Δ* respecto al WT (Whitehouse *et al.*, 2007). La segunda lista está compuesta por genes en los que se ha detectado la presencia *in vivo* del remodelador mediante la técnica de ChIP (Yen *et al.*, 2012). Siguiendo las conclusiones generales de los citados trabajos, en adelante al grupo de genes de la primera lista los denominaremos como genes *dependientes* de Isw2, y a los de la segunda lista como genes *unidos* por Isw2.

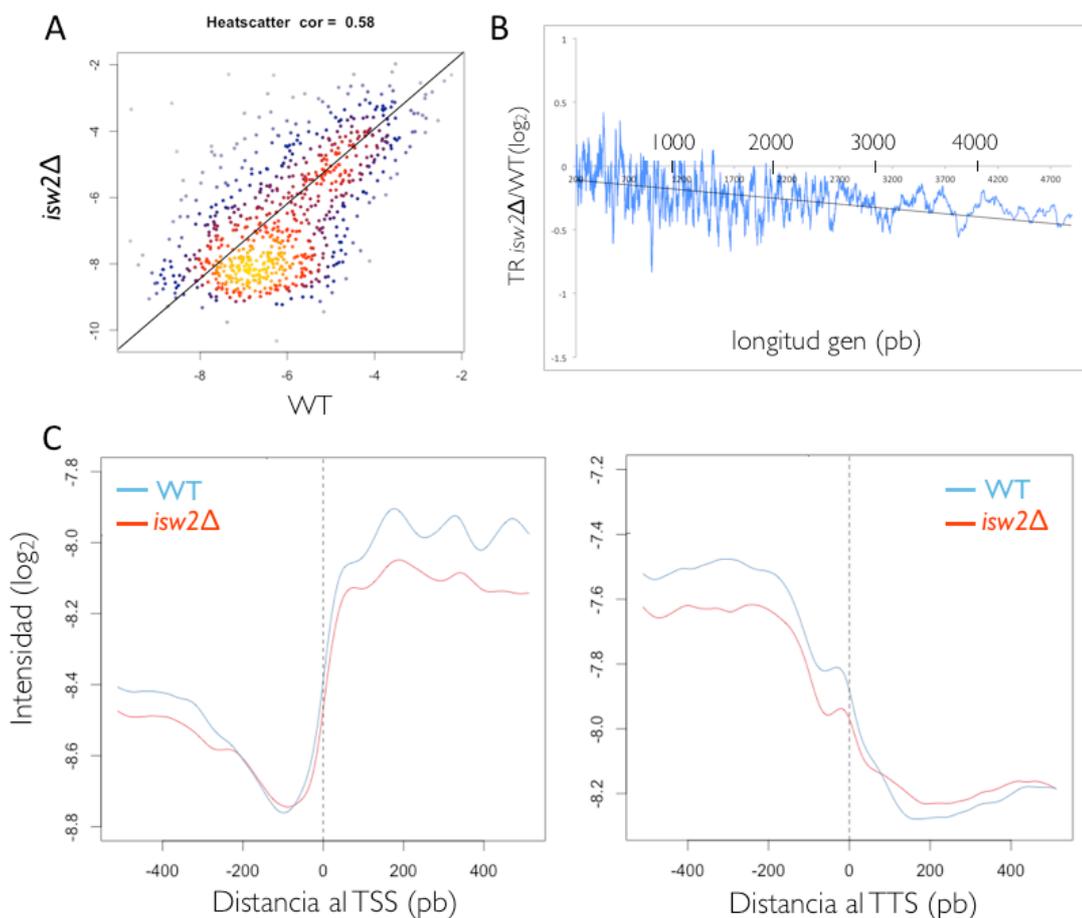


Figura 27. Caracterización de la cepa *isw2Δ* mediante BioGRO. (A) Correlación entre las TRs de la cepa silvestre y la cepa mutante. **(B)** Relación de los valores de nTR con respecto al WT, en relación a la longitud de los genes. **(C)** Comparación de los perfiles de RNAPs activas alrededor de los sitios de inicio y final de la transcripción.

El perfil de BioGRO resultante para cada lista de genes ofreció resultados aparentemente contradictorios: mientras que para los genes *dependientes* de *Isw2* no se observaron cambios en el posicionamiento de las RNAPs activas (figura 28, A), en los genes cuyos nucleosomas tienen *unido* *Isw2* se ve un desplazamiento hacia 3' de la magnitud esperada con respecto al WT (figura 28, B). Esta discrepancia en los resultados, *a priori* sorprendente, se debe al escaso solapamiento entre las dos listas, de únicamente el 5% (poco significativo estadísticamente, con p-valor <0.12) (figura 28, C). Esta diferencia entre los genes encontrados por los dos estudios invita además a pensar que el mecanismo de actuación de *Isw2*, y el impacto que éste tiene sobre los genes, posee una complejidad mayor de lo que en principio podría esperarse.

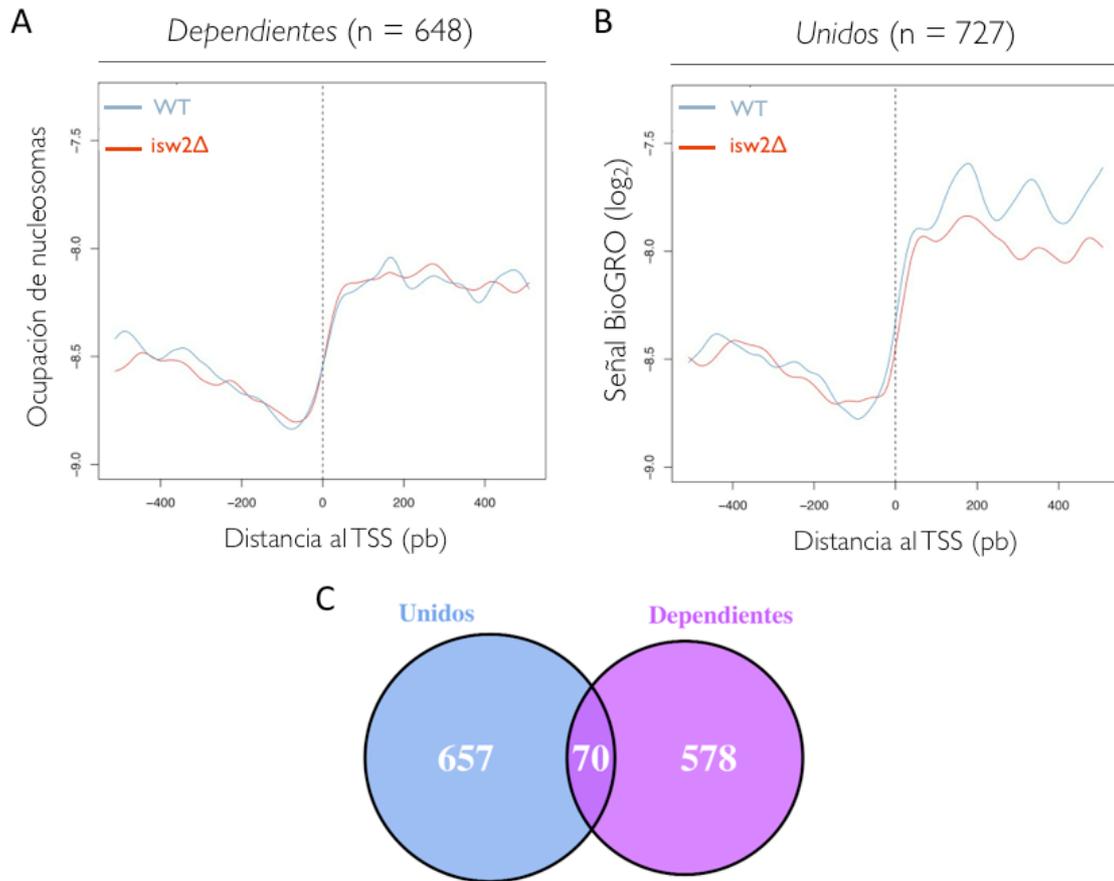


Figura 28. Transcripción naciente promedio de genes influenciados por Isw2 según estudios previos a esta tesis. (A) Perfil de los genes *dependientes* de Isw2 (de acuerdo con Whitehouse *et al.*, 2007) **(B)** Perfil de los genes *unidos* por Isw2 (de acuerdo con Yen *et al.*, 2007). **(C)** Solapamiento entre los grupos de genes de referencia publicados.

A pesar de ello, decidimos reanalizar los datos de posicionamiento de nucleosomas de Whitehouse y col. con nuestros propios métodos bioinformáticos con el fin de encontrar nuevos grupos de genes en los que apareciera también un cambio en el perfil de BioGRO. Para ello tomamos los datos crudos, alojados en el repositorio de datos experimentales de *microarrays* GEO, <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8813> y realizamos un agrupamiento de genes atendiendo al reposicionamiento de nucleosomas entre el WT y el mutante.

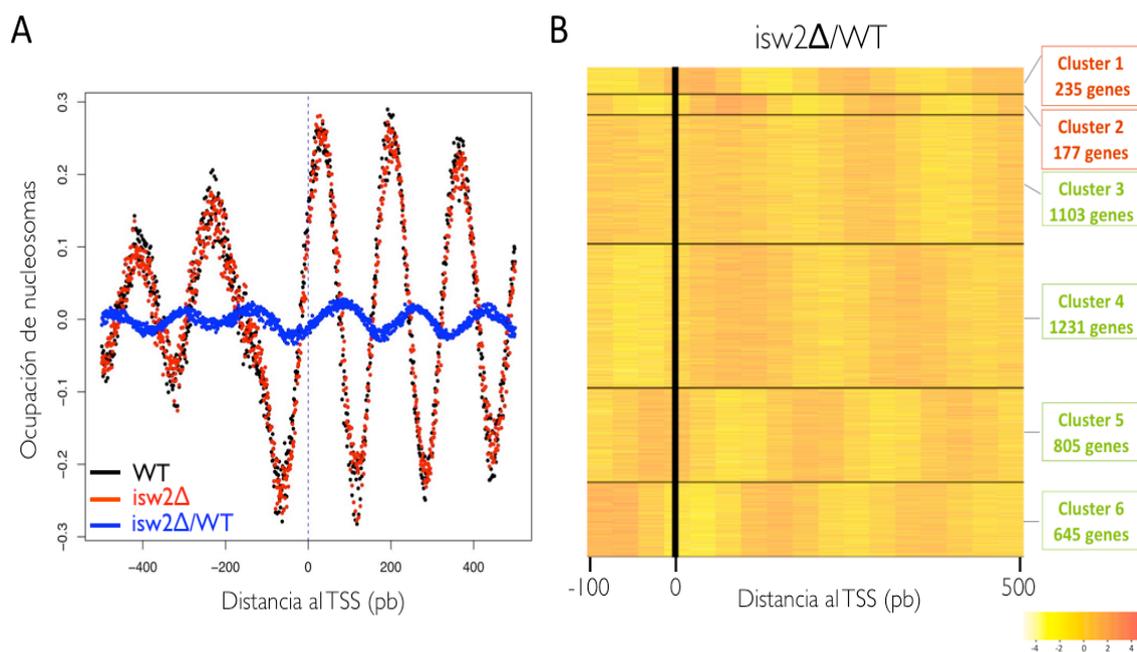


Figura 29. Estudio comparativo del posicionamiento de nucleosomas alrededor del TSS entre la cepa WT y el mutante *isw2Δ*. (A) Posicionamiento de nucleosomas a escala genómica alineados por el TSS y perfil resultante de la relación entre los datos del mutante/WT (trazo azul, representando el global del genoma). (B) mapa de calor resultante del agrupamiento de genes por el desplazamiento de nucleosomas entre una cepa WT y mutante. Código de color en unidades arbitrarias de intensidad (\log_2).

Para llevar a cabo el análisis del metagen 5' (figura 29, A) se alinearon los datos de densidad de señal respecto al TSS de todos los genes y se generaron tres perfiles dentro del gráfico: uno para los nucleosomas de la cepa WT (trazo de puntos negros), otro para los de la cepa mutante (puntos rojos), y otro para la relación entre los dos anteriores (puntos azules). De esta manera, la amplitud de la onda azul nos ayuda a valorar visualmente el grado de desplazamiento que presentan las posiciones promedio de los nucleosomas en cada cepa. Una línea plana indicaría que no hay variación, y la altura de la ondulación es proporcional a la discrepancia entre mutante y WT. Tal y como se ve en el gráfico, los perfiles globales de posicionamiento de nucleosomas generan una onda azul que no es completamente plana, lo cual indica que el mutante posee un ligero desplazamiento en la posición de sus nucleosomas

respecto a la cepa WT. Para encontrar el grupo o grupos de genes en los que es máximo este desplazamiento, llevamos a cabo un análisis de agrupamiento mediante k-medias con los datos de la relación *isw2Δ*/WT. El análisis dividió los genes en seis grupos distintos según su grado de desplazamiento (figura 29, B). A partir del análisis de k-medias se extrajeron las listas de genes de cada grupo y se generaron los correspondientes perfiles de BioGRO. De los seis, los grupos 1 y 2 fueron los que presentaron un desplazamiento más acusado. De ellos, el grupo 1 se descartó ya que el perfil de posicionamiento de nucleosomas situaba la presencia de un nucleosoma a unos ~100 pb antes del TSS en la cepa WT (no mostrado), con lo que los datos resultan difíciles de explicar, al mezclarse potencialmente los efectos de un posicionamiento no convencional en el WT, con un posible efecto de la ausencia de *Isw2*.

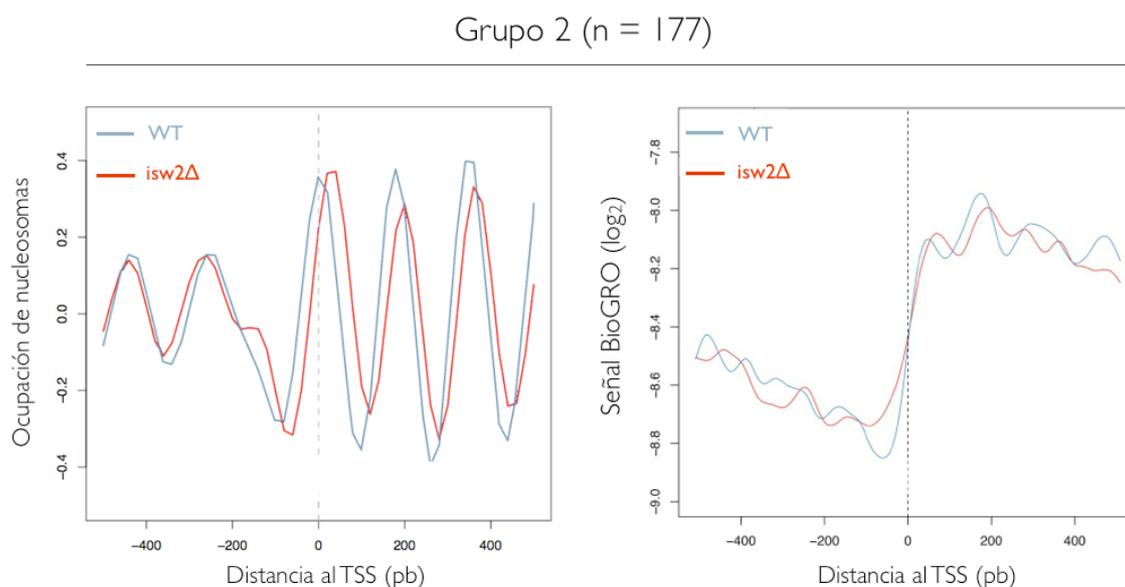


Figura 30. Análisis del cambio de posición de nucleosomas y RNAPs activas en el mutante *isw2Δ*. La figura de la izquierda muestra el perfil de nucleosomas alineados por el TSS para el grupo de genes del grupo 2 obtenido por k-medias (en figura 26, B). La figura de la derecha es el correspondiente perfil de BioGRO.

Por tanto, analizamos el posible desplazamiento del perfil de BioGRO sólo en el grupo 2 (177 genes). Como se observa en la figura 30 (derecha) el perfil de BioGRO, especialmente en los dos primeros picos, se encuentra también desplazado hacia la

derecha, de forma similar al de nucleosomas. Este desplazamiento entra dentro del rango promedio descrito en la literatura (~10 pb).

El análisis de solapamiento entre los genes del clúster 2 de nuestro análisis con el total de los genes encontrados en los dos estudios publicados (figura 31) pone de manifiesto la poca coincidencia de cada una de las estrategias de búsqueda y análisis elegida.

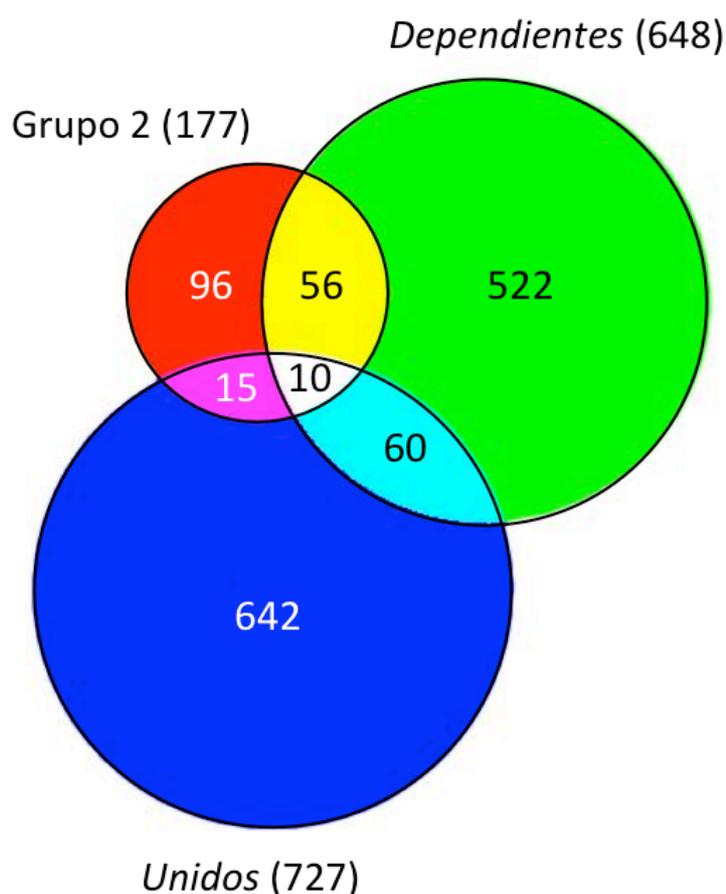


Figura 31. Solapamiento entre distintos conjuntos de datos analizados. El diagrama de Venn muestra los genes comunes entre nuestro análisis (círculo rojo), genes con *Isw2p* unido (círculo azul, Yen *et al.*, 2012), y genes dependientes de *isw2Δ* (círculo verde, Whitehouse *et al.*, 2007).

Estos resultados apoyan la hipótesis de que el patrón fluctuante en la señal de BioGRO visible al principio de los genes se genera como consecuencia de los cambios de

velocidad de la RNAPII debidos a la presencia de los nucleosomas, que constituyen un obstáculo para el paso de la maquinaria transcripcional a lo largo del cuerpo de los genes.

3.3-Análisis de la cepa mutante *dst1Δ*.

La naturaleza discontinua del paso de las RNAPs a través de los nucleosomas, en concreto la frecuencia y duración de las pausas de la maquinaria transcripcional sobre los genes, ha sido considerado por algunos como un mecanismo de regulación del que hacen uso las células para conseguir obtener diferentes tasas de transcripción para diferentes genes o grupos de genes, lo cual contribuye tanto a modular el grado de ruido de diferentes procesos celulares, como a incrementar la variabilidad entre las células de una población (Voliotis *et al.*, 2008).

Un suceso de pausa comienza cuando una RNAPs transcribe una región en la que la estabilidad del híbrido DNA-RNA es baja, lo que disminuye su velocidad de translocación, pudiendo llegar a detenerse e incluso moverse de manera retrógrada, en un proceso conocido como *backtracking* (Voliotis *et al.*, 2009, Cheung y Cramer, 2011, Gomez-Herreros *et al.*, 2012). Este movimiento hacia detrás hace que el RNA naciente deje de estar alineado en el sitio activo del holoenzima y la transcripción entra en un estado de parada, de duración variable (con un promedio *in vitro* de 8.1 seg, según Hodges *et al.*, (2009)). Ha sido descrito que las pausas de la maquinaria transcripcional no deben ser permanentes, ya que de ser así aumenta la probabilidad de que sean etiquetadas mediante ubiquitinación para su subsiguiente degradación por el proteasoma (Bregman *et al.*, 1996, Somesh *et al.*, 2005, 2007; Selth *et al.*, 2010). Las RNAPs salen de este estado de dos maneras principalmente, resumidas en la tabla 4: 1) por la actuación de las propias RNAPs, o 2) mediante la intervención de factores externos. Los mecanismos por los que las propias RNAPs salen de parada incluyen la realineación espontánea del transcrito naciente en el sitio activo gracias al movimiento anterógrado espontáneo de las RNAPs sobre el DNA molde (Hodges *et al.*, 2009), y/o su capacidad intrínseca para producir un corte endonucleolítico de la porción desalineada del RNA naciente. Esta actividad

enzimática es capaz de llevarla a cabo la propia RNAP, pero a muy baja tasa, con lo que normalmente es estimulada por el factor de elongación TFIIIS/Dst1 (Cheung y Cramer, 2011). La colisión con otra RNAP más retrasada en transcripción activa también se ha postulado como mecanismo de reactivación (Saeki y Svejstrup, 2009; Jin *et al.*, 2010). El otro factor externo que interviene es el complejo Ccr4-Not, el cual es el encargado de estabilizar el movimiento anterógrado espontáneo de las RNAPs sobre el DNA molde, impidiendo que vuelvan a moverse aguas arriba, promoviendo así el realineamiento (Kruk *et al.*, 2011).

Mecanismo	Factor que interviene	Referencia
Fluctuaciones anterógradas sobre el DNA molde	-Movimientos de las propias RNAPs -Complejo Ccr4-Not (estabiliza el movimiento) -Colisiones de las RNAPs en transcripción con RNAPs paradas	-Hodges <i>et al.</i> , 2009 -Kruk <i>et al.</i> 2011 -Jing <i>et al.</i> , 2011
Corte endonucleolítico de la porción desalineada del RNA naciente	-Actividad catalítica intrínseca de RNAPs -TFIIIS (estimula el corte)	Cheung y Cramer, 2011

Tabla 4. Resumen de los mecanismos involucrados en la reactivación de RNAPs paradas.

El mecanismo en el que nos centramos en esta tesis doctoral fue en el de reactivación enzimática de la RNAPs mediante el mecanismo del corte endonucleolítico estimulado por el factor TFIIIS/Dst1.

Para ello, decidimos ver si la ausencia de TFIIIS/Dst1 resultaba en un aumento de la tasa de *backtrack* de las RNAPs sobre los genes, que se reflejara fenotípicamente en el perfil global de distribución de RNAPs activas de BioGRO. Para ello aplicamos el método a la cepa *dst1Δ*, carente del factor TFIIIS/Dst1.

En la figura 32 se muestran los resultados de la caracterización del mutante, entre los que cabe destacar la bajada del nivel de TR naciente a nivel genómico respecto a la cepa WT (panel A), y un fenotipo de defectos en elongación (panel B) que se asemeja al

encontrado al caracterizar el mutante *rpb4Δ* (figura 23, B). La bajada del nivel de TR naciente a nivel genómico indica una reducción en la tasa de inicio de transcripción. Un efecto directo de bajada de inicio de transcripción en este mutante se ha descrito en el gen *GAL1* (Prather *et al.*, 2005). Por otra parte, de manera similar a lo observado en el mutante *rpb4Δ* (figura 23, B), se aprecia una bajada rápida de la relación de TR entre el mutante *dst1Δ* y el WT conforme aumenta la longitud de los genes (resaltado con la línea de tendencia local marcada en rojo en la primera Kb), lo cual sugiere que la presencia del factor TFIIIS/Dst1 es más necesaria cuanto más largo es el gen a transcribir.

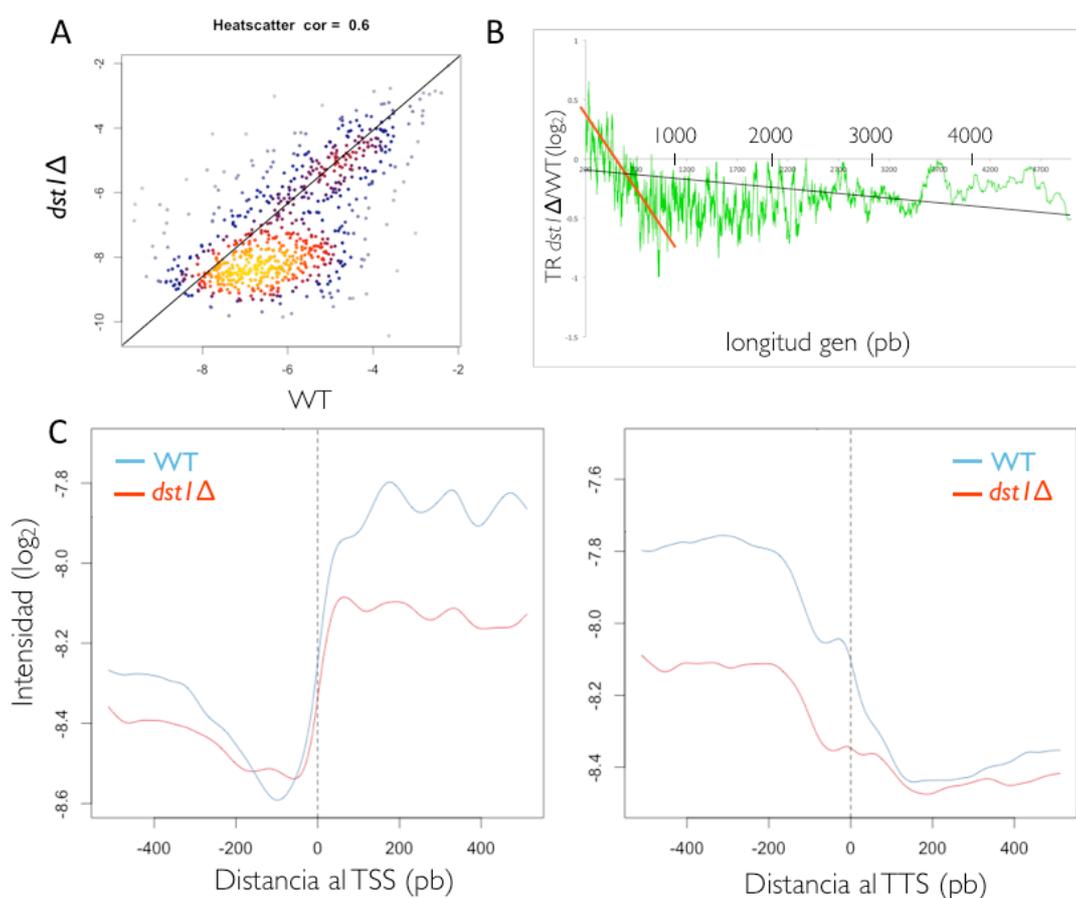


Figura 32. Caracterización de la cepa *dst1Δ* mediante BioGRO. (A) Correlación entre las TRs de la cepa silvestre y la cepa mutante. **(B)** Relación de los valores de nTR con respecto al WT, en relación a la longitud de los genes. **(C)** Comparación de los perfiles de RNAPs activas alrededor de los sitios de inicio y final de la transcripción.

Este efecto se ve más claramente al dibujar el perfil de BioGRO para grupos concretos de genes. Es el caso de los genes RP, conocidos por su alta dependencia del factor

TFIIS/Dst1 para su correcta transcripción (Gomez-Herreros *et al.*, 2012). Para dibujar el perfil se amplió la región a 1 kb alrededor del TSS, debido a que de esta forma se puede ver en un mismo gráfico toda la extensión de la ORF promedio de los genes RP (980 pb). La comparación (figura 33, A) revela como las señales discurren en paralelo a lo largo de la ORF excepto en la zona 5', donde la señal del mutante cae más rápidamente que la del WT (flecha en figura 33, A). Al igual que ocurría para el conjunto de genes en el mutante *rpb4Δ* (capítulo 2, figura 23, C), esta caída en la región 5' de los genes RP puede ser interpretada de dos maneras: como un aumento en la velocidad de las RNAPs, o una pérdida de señal de *run-on* a consecuencia de un aumento de *backtrack* y/o DO.

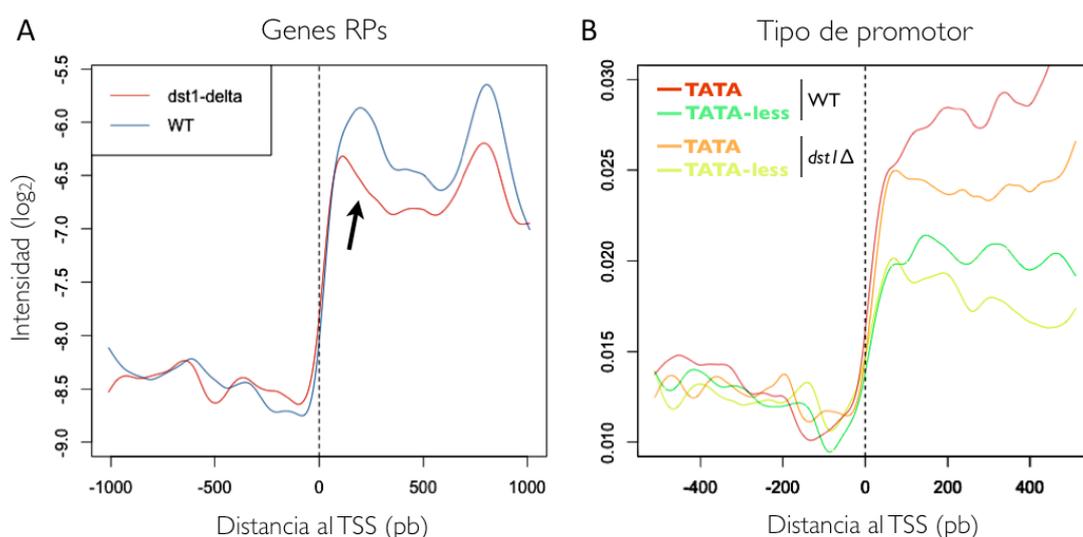


Figura 33. Defectos en la elongación del mutante *dst1Δ* en grupos concretos de genes. (A) La flecha del interior del gráfico señala el fenotipo de defectos en elongación en la zona inicial de los genes RP que exhibe el mutante. (B) Comparación de los perfiles de las dos cepas en genes clasificados según su tipo de promotor (con o sin caja TATA canónica).

Asimismo, también ha sido descrito (Gomez-Herreros *et al.*, 2012) que el factor TFIIS/Dst1 es importante para la correcta transcripción de la cromatina en los genes cuyo promotor posee una caja TATA. Al dibujar el perfil de los genes TATA y no-TATA desde el TSS para las dos cepas (figura 33, B) se observa claramente la caída más

nítidamente en los genes con promotores con caja TATA (trazo naranja), que en los sin TATA (trazo verde), confirmando los hallazgos previos.

En conjunto, los resultados nos permiten confirmar que la ausencia del factor de reactivación de la transcripción TFIIIS/Dst1 es importante para la elongación de las RNAPs. Siendo esta carencia especialmente perjudicial para la zona inicial de grupos concretos de genes: genes largos, RPs y genes cuyos promotores contengan caja TATA.

Capítulo 4. Caracterización del nascentoma de la RNAPIII.

4.1-Visión general de la transcripción naciente de la RNAPIII.

El *run-on* es una técnica que detecta transcripción proveniente de las tres RNAPs nucleares, siendo imposible discriminar, en principio, la contribución individual de cada RNAP al conjunto de la señal. Una manera clásica de poder separar la actividad de las distintas RNAP consiste en el uso de α -amanitina, una micotoxina peptídica que ejerce un efecto inhibitorio sobre ellas de manera diferencial. Así, la RNAPI es mayoritariamente insensible, la RNAPII es muy sensible y la RNAPIII muestra una sensibilidad intermedia (Rudd y Lusse, 1996, Dieci *et al.*, 2007).

En esta tesis se llevó a cabo un experimento de inhibición por α -amanitina, empleada a una concentración tal que eliminara la contribución de RNAPII, sin afectar a las otras dos. Para ello se realizaron dos experimentos de BioGRO en paralelo, siendo uno de ellos incubado durante 5 minutos con 50 μ M de α -amanitina, mientras que el otro, sin adición del péptido, se usó como control. Tras ello se hibridaron los RNAs nacientes en los microchips de embaldosado.

Al representar la intensidad de señal promedio para cada gen de cada experimento (figura 34, A) se confirma que los genes presumiblemente transcritos por la RNAPII (puntos negros) disminuyen considerablemente su señal, mientras que el grupo mayoritario de genes de la RNAPIII, los tRNAs (puntos rojos, $n = 274$), además de tener en general un nivel de señal mucho mayor al promedio de los genes de RNAPII (~50 veces), permanecen ajustados a la línea de tendencia que marca la diagonal del gráfico, lo cual indica que su señal no ha sido sensiblemente alterada.

Por otro lado, al comparar las señales de las regiones antisentido de los mismos genes (figura 34, B), encontramos dos aspectos interesantes: 1) a pesar del bajo nivel global de señal naciente antisentido capturada por BioGRO, una proporción de ésta se encuentra por encima del nivel de detección del método, ya que la desviación de la línea de tendencia respecto a la diagonal indica que la transcripción antisentido de

RNAPII es inhibible por α -amanitina. 2) la transcripción antisentido de los genes tRNA es muy baja, menor que del promedio de los genes de RNAPII, dentro de la zona de ruido técnico del sistema. Esto indica que los genes tRNAs tienen la transcripción antisentido bloqueada, a consecuencia probablemente de el nivel tan alto de transcripción sentido por la RNAPIII.

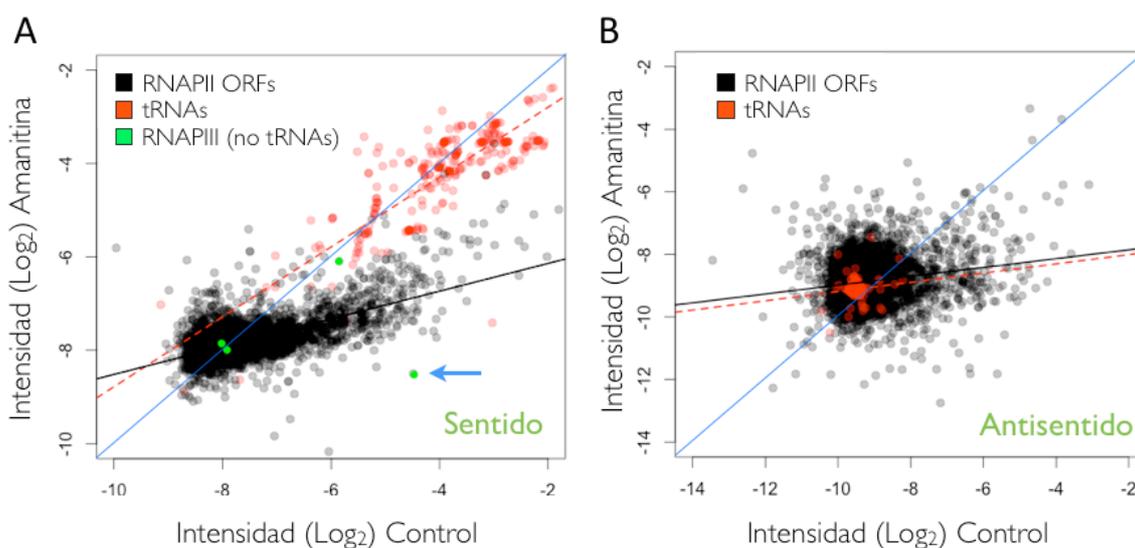


Figura 34. Estudio del cambio de señal de BioGRO tras un ensayo de inhibición selectiva de RNAPII con α -amanitina. (A) gráfico de comparación entre las señales de BioGRO en presencia o ausencia de α -amanitina. En negro se muestran los genes de la RNAPII y su línea de tendencia. En rojo se resaltan los genes tRNA, codificados por la RNAPIII, y su línea de tendencia. En verde se resaltan los 4 de los transcritos de RNAPIII que no pertenecen a tRNAs. **(B)** La misma comparación, pero tomando las señales de las regiones antisentido de cada gen. Las líneas azules marcan la diagonal del gráfico. La flecha azul señala la posición del gen snR6.

En relación a la TR de los otros genes transcritos por la RNAPIII aparte de los tRNAs, todos se muestran insensibles a la α -amanitina, con la notable excepción del gen snR6, cuyo producto es un RNA estructural del espliceosoma (flecha en figura 34, A).

En el análisis del metagen 5' de los genes de RNAPII tras el tratamiento con la amanitina se aprecia claramente como el perfil de densidad promedio decrece

sensiblemente comparado al no tratado (figura 35, A). El hecho de que la señal disminuya pero no desaparezca completamente, apreciable en nuestros resultados tanto a nivel de genes individuales (figura 34), como a nivel de perfil de metagen (figura 35, A), concuerda con el conocimiento previo que se tiene acerca de cómo la toxina reduce la tasa de translocación de la RNAPII, pero no es capaz de bloquearla completamente (Rudd y Luse, 1996, Antúnez y Pérez-Ortín, no publicado). El perfil promedio de las regiones de los tRNAs aparece en forma de un fuerte pico de señal localizado entre nucleosomas posicionados (figura 35, B). La señal promedio es ~50 veces mayor que la señal promedio de los genes de la RNAPII. Asimismo, llama la atención la rápida bajada de señal desde el centro hacia el final de los genes, que sugiere un aumento de la velocidad de las RNAPs en ese tramo.

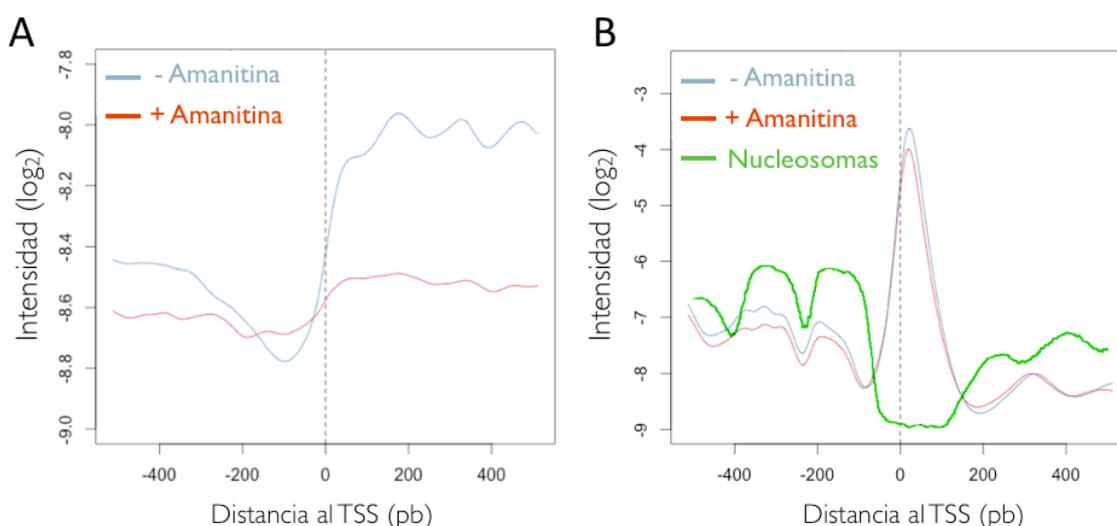


Figura 35. Análisis del metagen 5' tras un ensayo de inhibición selectiva de RNAPII con α -amanitina. A) Metagen 5' de genes con ORF. B) Metagen 5' de genes tRNAs. Nucleosomas: perfil de posicionamiento de nucleosomas alrededor de los genes de los tRNAs (tomado de Brogaard y Widom, 2012).

Por otra parte, cuando se analiza el perfil de transcripción de los tRNAs a partir de un ensayo de transcriptómica convencional (extracción, amplificación y marcaje de RNA total), se observa un perfil de transcripción muy débil y ruidoso (figura 36, A). Esto viene a confirmar los hallazgos previos de que los tRNAs, debido a sus estables

estructuras secundarias y sus extensas modificaciones post-transcripcionales, son pobres sustratos para la transcriptasa inversa (Dittmar *et al.*, 2006). La dificultad, por lo tanto, de tener una visión global de la transcriptómica de los genes de RNAPIII mediante otras técnicas ómicas indica que el método BioGRO es el mejor método para el estudio de la expresión global de esta clase de genes.

4.2- Estudio de los tRNAs con intrones.

Una de las características de los tRNAs es la presencia de regiones intrónicas en la secuencia de 60 de los 274 genes de *S. cerevisiae*. Los intrones de los tRNAs son cortos (25 nt en promedio), y se localizan a una distancia promedio de 37 nt del inicio del gen. Al analizar los tRNAs con intrones por separado, se ve que la densidad de RNAPs elongantes presenta un perfil distinto al del conjunto de los genes (figura 36, A y B), de manera que aparece un segundo pico de señal, más hacia 3' que el común a todos los genes, y que llega a alcanzar una altura parecida a éste. Este pico de señal indica la presencia una segunda zona de acumulación de RNAPIII dentro de los genes tRNA con intrones, que coincide aproximadamente con la zona de unión intrón-exón 3'.

Dado que el procesado de intrones no es co-transcripcional en estos genes (Yoshihisa *et al.*, 2003), no cabe esperar que el segundo pico de acumulación de RNAPs se deba al mismo mecanismo causante de la ralentización que se observa en los genes ORF con intrones (capítulo 2, fig 16, C). No obstante, ha sido descrito (Kraeva *et al.*, 2007) que las regiones más ricas en A+T afectan negativamente la estabilidad, y por tanto a la velocidad, del complejo de elongación.

Por tanto, decidimos investigar la existencia de un efecto de secuencia de esa región que pudiera estar afectando a la velocidad de la RNAPIII. Para ello calculamos el porcentaje de G+C promedio en tres regiones de cada gen: primer exón (+10 a +30 nt desde el inicio), intrón (+40 a +60 nt desde el inicio), y segundo exón (+69 a +90 nt desde el inicio). El gráfico de barras (figura 36, C) muestra como la región del intrón es significativamente menos rica en G+C que las regiones exónicas, y que la región

equivalente en los genes sin intrones (datos no mostrados). Este resultado sugiere que, en los tRNAs con intrones, las RNAPs sufren un descenso de su velocidad al transcribir la zona del intrón, rica en A+T, y que después la velocidad se recupera hasta el final del transcrito.

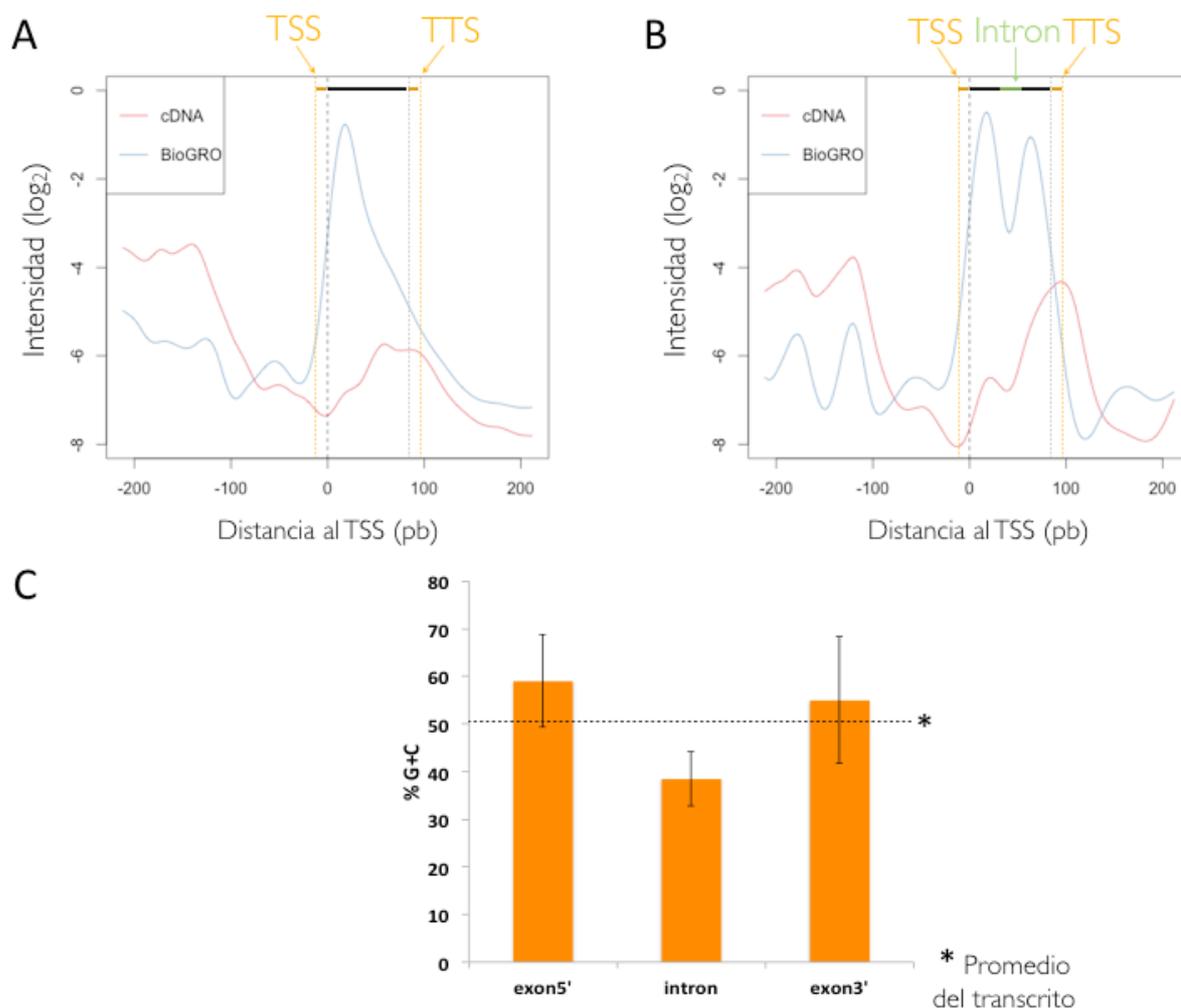


Figura 36. Estudio de la transcripción de los genes tRNA con intrones. A) Perfil promedio de transcripción de todos los genes tRNA obtenido mediante BioGRO o partiendo de RNA total. **B)** Perfil promedio de transcripción de los genes tRNA con intrones. Las anotaciones en la parte superior de ambos gráficos hacen referencia a la estructura del pre-tRNA: extensiones 5' y 3' en naranja, intrones en verde. **C)** Contenido en G+C a lo largo de la secuencia de DNA de los tRNAs con intrones.

4.3- Selección traduccional y transcripción naciente.

La traducción de los mRNAs en eucariotas es un proceso regulado en distintas etapas y por múltiples mecanismos (Ingolia *et al.*, 2009). Uno de los sucesos que determinan la cantidad de una determinada proteína es su tasa de síntesis, que viene determinada en gran medida por la disponibilidad de la maquinaria de traducción. Uno de los componentes más importantes de dicha maquinaria son los tRNAs. La levadura *S. cerevisiae* contiene 274 genes de tRNAs nucleares, transcritos por la RNAPIII, con un alto grado de similitud de secuencia. En base a la secuencia del triplete anticodón, los diferentes tRNAs han sido clasificados en especies o familias de genes según el aminoácido (aa) que portan (isogenes). Así, los 20 aas, descodificados por los 274 tRNAs, se agrupan en 42 familias, en las que cada familia contiene entre 1 y 16 copias del mismo isogen. Se ha descrito que no todas las especies de tRNA se encuentran en la misma proporción en la célula, siendo algunas más abundantes, y otras más raras (Ikemura, 1985). Las especies más abundantes corresponden a los aas que se encuentran en una proporción mayor en las proteínas, como consecuencia de un proceso evolutivo por el que la abundancia relativa de los tRNAs se ha co-adaptado a esa abundancia de aas. Además, dentro de cada grupo de codones codificantes de un único aa, se ha encontrado un sesgo hacia ciertos codones leídos por un mismo tRNA isoceptor en detrimento de los otros codones sinónimos (y de sus tRNA lectores). Este sesgo en el uso de codones ha evolucionado como una manera de optimizar el proceso de la traducción, de manera que en organismos de crecimiento rápido como la bacteria *E. coli*, o la levadura *S. cerevisiae*, el uso de subconjuntos de codones óptimos (también denominados como *preferidos*, o *favoritos*, Sharp y Li, 1987) es capaz de conseguir tasas de traducción más altas. La fuerza evolutiva responsable de esto se conoce como selección traduccional (Percudani *et al.*, 1997, dos Reis *et al.*, 2004).

A pesar de que la existencia de selección traduccional es un mecanismo ampliamente aceptado, la manera que tiene una célula de conseguir los determinados niveles de cada tRNA no se conoce tan bien. Estudios previos han encontrado que existe una relación directa entre el número de copias de genes de cada especie de tRNA en el genoma y la abundancia del tRNA correspondiente en la célula (Percudani *et al.*, 1997,

Tuller *et al.*, 2010), lo cual les lleva a la conclusión de que la transcripción de cada isogen permanece constante, y que la cantidad de tRNA final en la célula se consigue mediante un incremento en el número de copias genómicas. A pesar de ello, hasta ahora no se han medido las tasas de transcripción de los genes de tRNAs, con lo que estas afirmaciones carecen de una demostración formal. Como se ha descrito anteriormente en esta tesis, la técnica de BioGRO sí es capaz de medir directamente la nTR de cada uno de los genes de tRNA de la levadura, con lo que se realizó un estudio detallado de los niveles de transcripción de esta clase de genes que nos permitiera responder a estas preguntas.

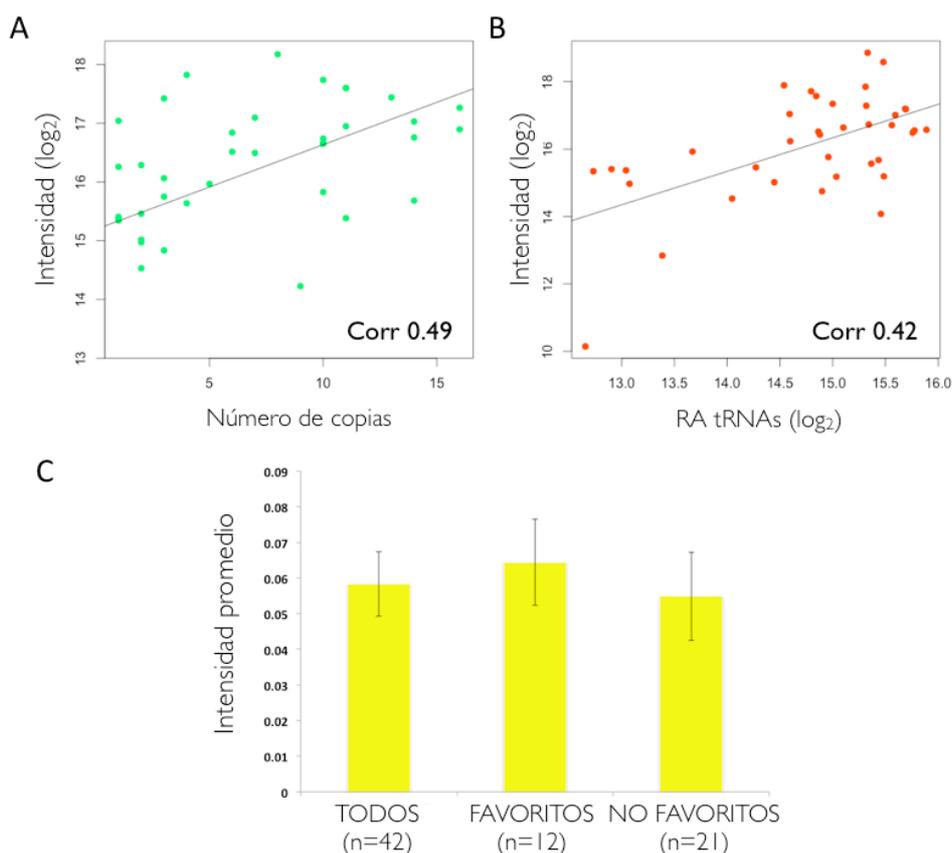


Figura 37. Estudio de los factores que caracterizan el nivel de expresión de los tRNAs. A)

Relación entre la nTR promedio medida por BioGRO de cada familia de tRNAs y el número de copias de los genes que componen cada familia. **B)** Correlación entre la nTR medida por BioGRO y la cantidad de tRNAs (RA, tomada de Tuller *et al.*, 2010). **C)** Nivel de nTR promedio de tRNAs según el uso de codones.

Para obtener un valor único de TR por familia de tRNA, se promedió el valor de TR individual de los isogenes correspondientes. La comparación de esos valores, tanto con el número de copias de cada familia, como con la abundancia de tRNAs en el citoplasma (obtenida mediante hibridación de los transcritos en microchips específicos de tRNA, Dittmar *et al.*, 2004, Tuller *et al.*, 2010), muestra una correlación positiva significativa (figura 37, A y B). A este respecto, y a pesar que existe cierta correlación, sorprende la ausencia de una relación más directa de proporcionalidad entre número de copias e intensidad de señal de hibridación obtenida, algo que no se refleja ni con nuestros datos ni con los datos publicados que comparan concentración de tRNAs citoplasmáticos con su número de copias en el genoma (Figura 1 de Tuller *et al.*, 2010).

4.4- El nascentoma de RNAPIII en las cepas mutantes *dst1Δ* e *isw2Δ*.

La transcripción de la RNAPII ha sido históricamente la más estudiada, al igual que el papel que ejercen sobre ella tanto los factores de transcripción como los remodeladores de cromatina. Así, son comparativamente escasos los estudios para caracterizar las proteínas involucradas en la transcripción de genes de RNAPIII. Entre ellos, existen algunos que le han asignado una función importante tanto al factor de transcripción TFIIS/Dst1 (Ghavi-Helm, 2008), como al remodelador de cromatina Isw2 (Gelbart *et al.*, 2005, Kumar y Bhargava, 2013). El primero de los estudios, basado en la IP del factor TFIIS/Dst1 a escala genómica, revela que la práctica totalidad de los genes de RNAPIII están unidos al factor, y que la ausencia de éste condiciona la expresión de unos pocos tRNAs, y únicamente en condiciones de crecimiento a baja temperatura (16°C). Estos hallazgos les llevan a proponer a TFIIS/Dst1 como un nuevo factor de transcripción general de RNAPIII. El segundo de los trabajos, basado en IP y RT-qPCR, encuentra al factor Isw2 unido a la región 5' NFR de los tRNAs, y que el doble mutante *isw1Δ/isw2Δ* presenta defectos tanto en el posicionamiento de nucleosomas cercanos al gen, como en la expresión de algunos tRNAs. A la vista de estos resultados, decidimos profundizar en este tema, caracterizando para ello la TR de RNAPIII en los mutantes *dst1Δ* y *isw2Δ*.

Para ello, se obtuvieron los valores de nTR de BioGRO de cada una de las 42 familias de tRNAs de las cepas mutantes de la misma manera que para el WT (ver apartado 4.2). En la matriz de gráficos de correlación (figura 38) se ve que la peor correlación corresponde a la comparación *isw2Δ*/WT (*Spearman* = 0.59), y la mejor correlación corresponde a la comparación *dst1Δ*/WT (*Spearman* = 0.72), quedando la comparación entre los mutantes *dst1Δ*/ *isw2Δ* con un valor intermedio (*Spearman* = 0.69).

El mapa de calor resultante de ordenar los genes de la cepa WT de mayor a menor TR y compararlos con los valores correspondientes de los genes de las cepas mutantes muestra una alteración general de los niveles de transcripción de los tRNAs en los mutantes, mucho mayor de lo que se había descrito hasta la fecha (figura 39). De esta manera, a *grosso modo*, existen dos grandes grupos de genes que muestran alteración en el patrón de transcripción naciente en ambos mutantes en sentido de bajada de la transcripción (señalados dentro de recuadros rojos en la figura), y otros cuya alteración es más específica de uno u otro, y en sentido de aumento de la nTR (recuadros verdes).

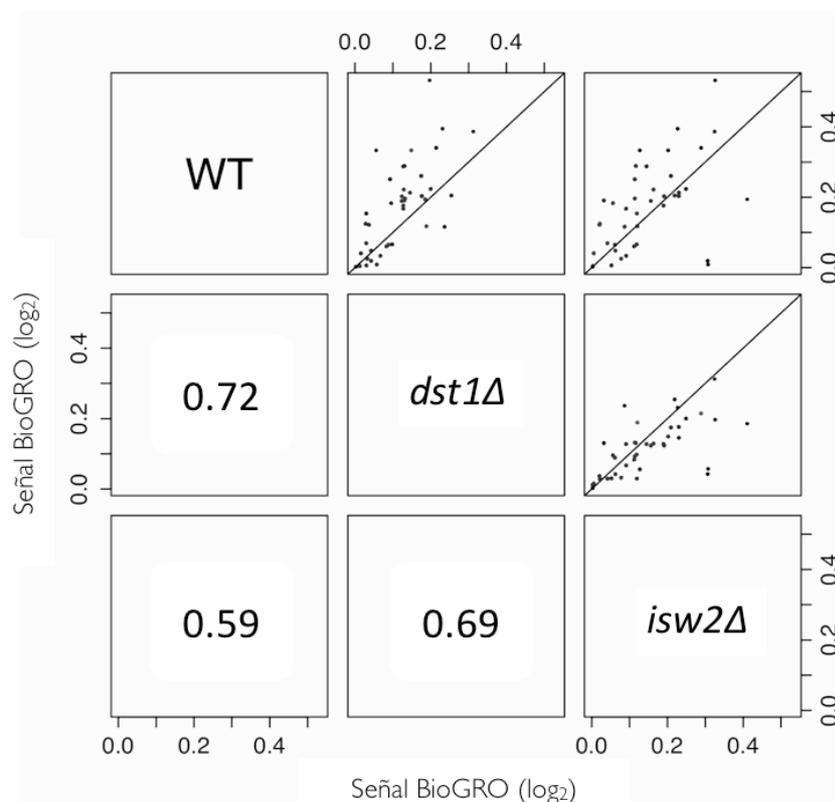


Figura 38. Comparación de la expresión de las familias de tRNAs en distintas cepas. Los números representan el valor de correlación de *Spearman* del gráfico correspondiente. Cada punto representa la media de nTR de una familia de genes de tRNA.

En relación al mutante *dst1Δ*, nuestros hallazgos demuestran que existe un efecto mayor en la transcripción de la RNAPIII que lo que se había visto anteriormente, incluso en condiciones de crecimiento en laboratorio óptimas (YPD, 30°C).

En cuanto al mutante *isw2Δ*, no fue posible hacer la comparación directa entre transcriptomas, ya que en el trabajo de Kumar *et al.* (2013) no se estudia el mutante simple *isw2Δ*, sino el doble mutante con *isw1Δ*. Además, en la discusión del trabajo se dice que los efectos en la alteración del patrón de posicionamiento nucleosomal son menores en el doble mutante que en los mutantes simples. Esto lo atribuyen a que probablemente se trata de remodeladores de cromatina con función antagónica actuando sobre un conjunto de genes similar, con lo que los efectos de su ausencia, tanto en posicionamiento nucleosomal, como presumiblemente en transcripción, se anularían. A pesar de ello, el solapamiento entre los genes con patrón alterado de expresión encontrado en los dos trabajos es mayor, encontrándose una coincidencia en 6 de las 8 familias (señaladas con un asterisco negro en la figura 39).

Cabe destacar también que los hallazgos del trabajo de Kumar *et al.* (2013), en los que se encuentra un patrón de posicionamiento nucleosomal en la cepa WT de los cuatro isogenes que codifican para el aa iniciador metionina ($\text{tRNA}_i^{\text{Met}}$), diferente al resto, con arquitecturas potencialmente represivas de la transcripción, no se traducen en un menor nivel de nTR medida por BioGRO (figura 39, asterisco rojo).

En conjunto, nuestros resultados aportan más datos a favor de que, tanto el factor de transcripción TFIIIS/Dst1, como el remodelador de cromatina Isw2, tienen un papel importante en la transcripción de los genes de tRNAs.

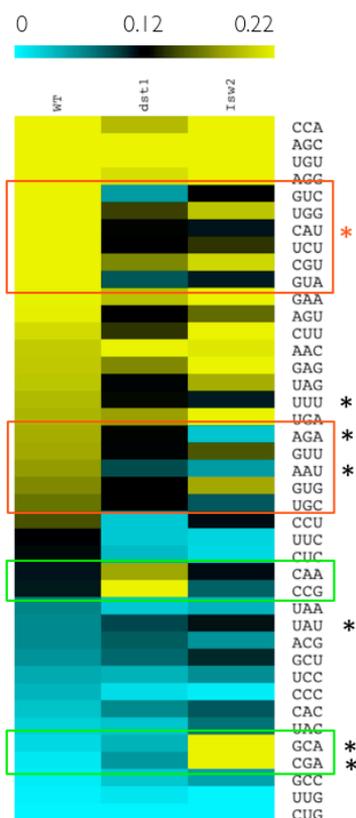


Figura 39. Efecto de la ausencia de Dst1 o Isw2 en la nTR de los genes tRNA. Cada fila corresponde al promedio de nTR de una familia de genes tRNAs. En verde se resaltan las familias cuya expresión resulta más afectada en uno de los dos mutantes, y en rojo las familias en las que ambos mutantes muestran cambios en la expresión. El asterisco rojo es el tRNA^{Met}, y los asteriscos negros son los genes diferencialmente expresados también en un doble mutante *isw1Δ/isw2Δ* (Kumar y Bhargava, 2013). Código de color en unidades arbitrarias de intensidad de señal (\log_2).

4.5-Los genes tRNA como puntos calientes de inserción de elementos transponibles.

Relación con la nTR.

El genoma de *S. cerevisiae* alberga cinco familias de retrotransposones tipo LTR (Ty1 a Ty5), estructural y funcionalmente relacionados con los retrovirus (Boeke and Sandmeyer, 1991). Se han documentado más de 300 inserciones de estos elementos, lo que en conjunto supone un 3% de las 12 MB del genoma en la cepa S288c (Kim *et al.*, 1998). La transcripción de estos elementos es dependiente de RNAPII, y en su proceso

de maduración se les añade caperuza en el extremo 5' y cola Poli(A) en 3'. A pesar de que la mayoría de ellos posee una TR baja (como resultado de promotores débiles e inserción en zonas heterocromáticas del genoma (Morillon *et al.*, 2002)), los mRNAs de los transposones de la familia Ty1 son de los mRNAs más abundantes de la célula, llegando a representar un 0.8% del RNA total de la levadura (Curcio *et al.*, 1990), lo cual se explica por su estabilidad, excepcionalmente alta (Curcio y Garfinkel, 1992).

Se ha descrito también que los transposones Ty tienen preferencia por insertarse en la vecindad de genes de la RNAPIII, sean tRNAs o no (Bolton y Boeke, 2003, Qi *et al.*, Mularoni *et al.*, Baller *et al.*, 2012). A pesar de ello, son pocos los trabajos que han estudiado una posible interacción entre las maquinarias de la RNAPII y la RNAPIII que resultara en una expresión diferencial de los genes de tRNAs. A nivel de unos pocos genes individuales, se ha descrito que no existe un gran efecto de la expresión de los transposones en la expresión de los tRNAs adyacentes (Kinsey y Sandmeyer, 1991, Bolton y Boeke, 2003). A pesar de ello, y dado que no existe un estudio a nivel genómico de esta relación, decidimos profundizar en ello con los datos de BioGRO.

Para comprobar si la transcripción naciente de los tRNAs con inserciones cercanas de retrotransposón era significativamente distinta de la de tRNAs sin inserción cercana, inspeccionamos el genoma para catalogar manualmente los tRNAs según tuvieran o no un retroelemento Ty a una distancia de 800 pb aguas arriba o aguas abajo (tanto en la misma cadena como en la cadena inversa). Una vez obtenida la lista, calculamos la TR naciente promedio de los genes dependiendo de la orientación del par tRNA-transposón. Los resultados (figura 40) muestran que más de la mitad de los tRNAs poseen un elemento transponible en su vecindad, que las orientaciones tándem y divergentes son mayoritarias, y que la TR promedio de los genes con y sin transposón no es significativamente distinta. Estos resultados están de acuerdo con un modelo en que la combinación de promotores débiles y estructura cromatínica represiva son los responsables de la baja expresión general de los elementos transponibles, más que una interacción competitiva entre las maquinarias de RNAPII y RNAPIII.

Por otra parte, se ha descrito que los transposones tienen preferencia por insertarse en regiones ricas en nucleosomas posicionados que flanquean genes, y que por tanto se benefician de la estructura nucleosomal particular de las regiones tDNA, que

obedece a tal estructura (ver figura 36, B, y Baller *et al.*, 2012). Esta estrecha relación entre los tRNAs y los transposones ha servido para hipotetizar que los genes de tRNA podrían dispersarse por el genoma gracias a la retrotransposición (Weiner *et al.*, 1986).

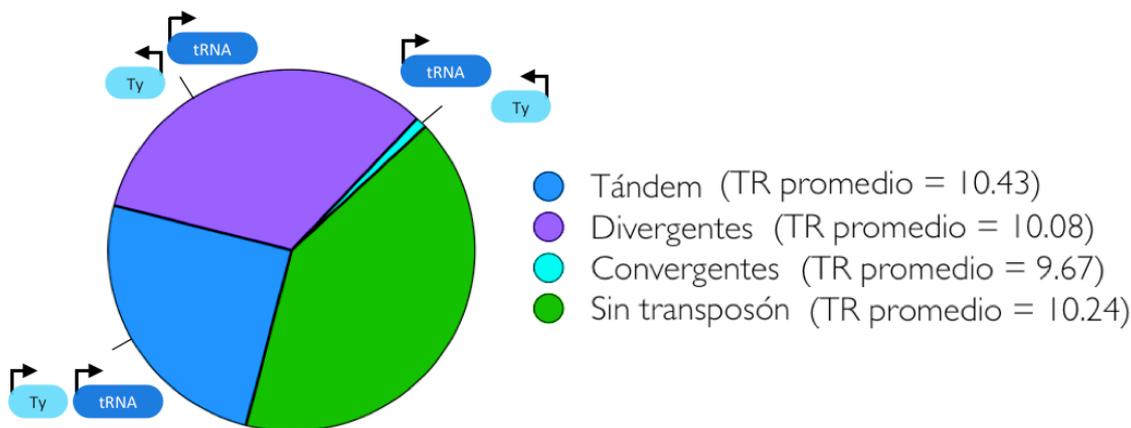


Figura 40. Estudio de la relación entre la expresión de los tRNAs respecto a la ocurrencia de elementos transponibles en sus proximidades, y sus diferentes orientaciones. Los valores entre paréntesis son el nivel de nTR promedio.

A pesar de ello, un estudio detallado de la relación entre las familias de tRNAs con mayor número de copias de isogenes y la aparición de un transposón en su vecindad muestra que no existe una asociación significativa entre los dos aspectos (figura 41).

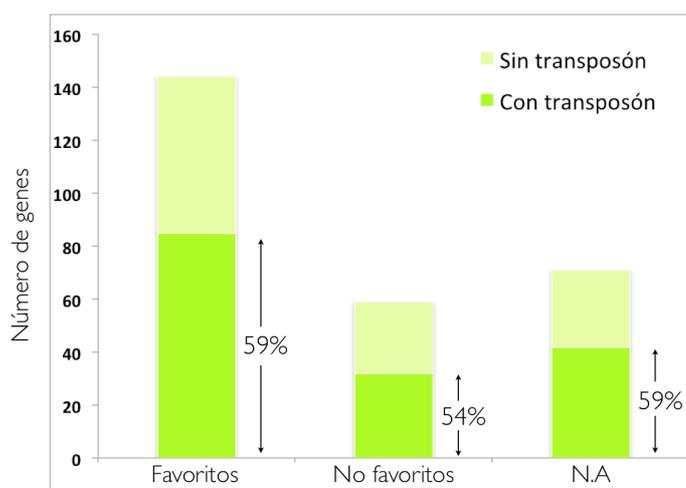


Figura 41. Análisis de la relación entre el sesgo en el uso de codones de los tRNAs respecto a la existencia de elementos transponibles en sus proximidades.

DISCUSIÓN GENERAL

La transcripción en eucariotas es un proceso complejo que requiere de la participación regulada de muchos factores de manera orquestada. El conocimiento sobre cada uno de los factores que intervienen en el proceso no ha hecho más que crecer cada año, en gran parte gracias a la puesta a punto de nuevas técnicas *ómicas* (o la actualización de técnicas pre-existentes), con las que se han obtenido cantidades ingentes de datos que han ayudado a arrojar luz sobre la organización y el funcionamiento de las diferentes etapas de la expresión génica. Gracias a la optimización y diversificación de técnicas de estudio global de la estructura y la regulación de los genomas, una de las áreas más intensas de avance científico en los últimos 10 años está siendo la genómica, tanto estructural (secuenciación y caracterización del DNA de muchas especies), como funcional (análisis de la expresión de esos genomas). El presente trabajo de tesis se ha enfocado hacia la revisión y optimización de una de las técnicas más útiles para estudiar la transcripción (el GRO radiactivo o *clásico*, desarrollado en nuestro laboratorio), con el objetivo primordial de reducir sus limitaciones, ampliar su alcance, y asegurar su capacidad de adaptación a las nuevas plataformas tecnológicas de estudios genómicos existentes.

El GRO sirve para la medición de la densidad de las RNAPs elongantes sobre los genes (la transcripción naciente) a escala genómica. Sin embargo, y tal y como se comenta en la introducción, no es la única técnica que aborda esta cuestión. Existen toda una serie de técnicas similares, que también suministran información sobre la transcripción naciente a escala global, pero con diferentes capacidades en cuanto a la detección de la actividad de las RNAPII. Su comparación con nuestros resultados puede servir para una mejor interpretación de los mecanismos empleados por las RNAPs durante la transcripción.

1. Validación de la autenticidad del RNA naciente.

A diferencia de los estudios de transcriptómica clásica, en los que la muestra de partida a interrogar es el contenido total de RNA de una célula (análisis *single-cell*, Tang *et al.*, 2009) o un conjunto de células, el estudio del transcriptoma naciente o *nascentoma* requiere del aislamiento y caracterización de las moléculas de RNA naciente, las cuales representan una fracción ínfima del RNA total de una célula (alrededor de un 0.34% según los cálculos de Churchman y Weissman, 2011). Dado que las moléculas de mRNA

sufren todo un proceso de cambio y maduración desde que se sintetizan a partir de su molde de DNA hasta que propagan su información al ser traducidas a proteínas en el citoplasma, es importante tener en cuenta que la caracterización del contenido de RNA total maduro de una célula no es informativa de las dinámicas de funcionamiento y regulación de las etapas iniciales de su síntesis. En ello radica la importancia de poder discriminar entre el estudio de los diferentes transcriptomas de la célula: el naciente, y el maduro. Técnicas como el GRO han sido diseñadas e implementadas en organismos como la levadura *S. cerevisiae* teniendo en cuenta esta consideración de partida, de manera que las conclusiones basadas en el análisis de sus resultados no estén sometidas a ambigüedades derivadas de no saber si se está caracterizando una fase u otra del ciclo de maduración de los mRNAs. Por otro lado, la aparición de nuevas tecnologías de análisis del transcriptoma de muy alta resolución y sensibilidad que suelen requerir de una amplificación previa (*tiling arrays* y secuenciación masiva), ha provocado que aumente la probabilidad de contaminar las muestras con RNA maduro, introduciendo así ruido en los datos y restando calidad a los resultados. A pesar de ello, dichas técnicas han hecho hincapié variable en los aspectos de control y validación de la autenticidad del material de estudio, lo cual es importante tener en cuenta a la hora de interpretar los resultados que de ellas se derivan.

El presente trabajo de tesis se ha centrado en la optimización del método *clásico* de GRO de *S. cerevisiae* (García-Martínez *et al.*, 2004) para sacar partido de la tecnología de alta resolución de los *microarrays* de embaldosado, haciendo especial hincapié en la autenticidad del RNA naciente de partida. Para ello se ha optado por un planteamiento metodológico similar al del GRO radiactivo, pero evitando aquellos pasos de tratamiento de las muestras más proclives a la introducción de sesgos técnicos. El GRO se basa en el marcaje radiactivo de los RNAs nacientes por las RNAPs transcripcionalmente activas. Una vez marcados esos RNAs, se extraen junto con el resto del contenido de RNAs de la célula y se hibridan directamente en macrochips de nailon. Estos macrochips están diseñados de manera que existe una sonda de DNA bicatenario correspondiente a la secuencia de la ORF de cada gen de la levadura. La ventaja fundamental de esta aproximación es que la señal radiactiva que se obtiene es inequívocamente informativa del nivel de transcripción naciente de cada gen, ya que los RNAs maduros no han sido marcados y por tanto son *invisibles*. A pesar de la baja proporción de RNA naciente (marcado) frente al RNA maduro

(no marcado), la hibridación de los macrochips con el total del RNA extraído de las células resulta estrictamente proporcional a la concentración de cada molécula de RNA naciente sin que la competencia de moléculas maduras no marcadas afecte de forma significativa (Hirayoshi y Lis, 1999). Ello se debe, probablemente, a que la concentración de RNA en la solución de hibridación es relativamente baja al ser el volumen de hibridación para esos chips bastante grande (3-5 mL) y la cantidad de sonda en cada punto de hibridación muy alta por el gran tamaño del punto de aplicación (1 mm de diámetro). Por el contrario, una de las limitaciones que posee el método es que se extrae un solo dato de intensidad de señal de hibridación promedio por cada gen, con lo que no se puede saber la posición física de las RNAPs sobre los genes y, al contener sondas únicamente de genes, tampoco de las regiones no génicas. La tecnología de *microarrays* de embaldosado de la casa comercial Affymetrix, que apareció posteriormente a los *macroarrays* de nailon, se basa en la hibridación de la muestra de partida sobre un soporte que contiene sondas monocatenarias de una longitud fija de 25 pb que cubren solapadamente todo el genoma de la levadura *S. cerevisiae*, en ambas hebras. Esto permite obtener muchos más datos de señal por cada región, y en ambas cadenas separadamente. Dado que se sabe que un 85% del genoma de la levadura se transcribe en condiciones normales de crecimiento en medio rico (David *et al.*, 2006), es importante no restringir la existencia de transcritos a regiones particulares que hayan sido previamente anotadas como genes. Los inconvenientes de partida fundamentales que encontramos al afrontar el uso de esta nueva técnica fueron: 1) que el método está basado en la fluorescencia, con lo que no es compatible con la metodología radiactiva tradicional, y 2) que la preparación de la muestra incluye etapas con un elevado potencial de incluir contaminación con RNA maduro no marcado (como amplificación mediante PCR, o transcripción *in vitro*). Esto último es debido a que, en estos microchips, la cantidad de sonda es muy reducida por las dimensiones del *spot* (11 μm de diámetro) y, sobre todo, por la elevada concentración de RNA en la solución de hibridación con la que se trabaja, que puede llegar a ser 100 veces mayor. Para solventar tales problemas decidimos, por una parte, probar marcajes alternativos al radiactivo, y por otra, hibridar la muestra directamente sobre el microchip, alterando el protocolo que se utiliza usualmente en Affymetrix para preparar las muestras. Experimentos previos en nuestro laboratorio habían permitido una hibridación directa de un experimento de GRO en el que se marcaba el RNA naciente con un nucleótido

biotinilado (Vicent Pelechano, tesis doctoral). No obstante, el análisis de los datos mostró que el nivel de señal obtenido con esta aproximación era excesivamente bajo, solamente distinguible del ruido de fondo en unos pocos genes de alta transcripción (fundamentalmente los de tRNAs y rRNAs). Como alternativa se probó a amplificar la señal original siguiendo los pasos del protocolo de preparación de la muestra que recomienda por defecto la casa comercial, pero los datos fueron poco concluyentes, ya que el perfil de la muestra de RNA naciente amplificada era prácticamente indistinguible de una hibridación de RNA maduro. Al comienzo de esta tesis se retomaron por tanto los experimentos encaminados a hacer un GRO no radiactivo e hibridarlo en un microchip de embaldosado. Visto que la señal obtenida al hibridar directamente una alícuota de la muestra de GRO era muy baja, se decidió digerir con RNasa A antes de llevar a cabo una reacción de *run-on* radiactivo, por las siguientes razones: 1) para poder enriquecer la muestra inicial en RNAs nacientes, 2) poder estimar la extensión de la elongación y su dependencia con la concentración de nucleótidos añadidos con los que ajustar las condiciones de reacción, y 3) ser capaces de poder aumentar la resolución del posicionamiento de RNAPs activas sobre el genoma de la levadura eliminando las porciones distales de los RNAs nacientes que son productos de transcripción previos a la reacción de *run-on*. Por una parte, tal y como muestran los resultados de las figuras 1-3, los RNAs maduros de las especies mayoritarias se degradan tras el tratamiento, quedando los RNAs nacientes marcados como un *smear* o *borrón* de fragmentos en la zona inferior de los geles, y, por otra parte, no existe impedimento posterior para que los nucleótidos radiactivos sean incorporados a los RNAs nacientes por *run-on*, al mismo nivel que en las células control sin digerir. Además, comprobamos que al congelar las células de manera rápida con nitrógeno líquido y mantenerlas después a -20°C durante unas horas antes de hacer el *run-on* se consigue un aumento de dos veces en el marcaje.

Tras estas optimizaciones iniciales, y a raíz de la aparición en la literatura de métodos de caracterización de transcritos nacientes que hacían uso de ellos, se testaron toda una batería de marcajes distintos, no radiactivos, para ver cual era el que daba un mayor rendimiento (sumarizados en figura 4). Estos marcajes tenían la ventaja adicional de permitir purificar selectivamente los RNAs marcados, disminuyendo así considerablemente el riesgo de contaminación con RNA maduro. De entre ellos, tanto el marcaje de los RNAs nacientes por *run-on* con Bromo-UTP como con Metil-CTP, y la

posterior purificación selectiva con anticuerpos u otras proteínas de unión específicas, no mostraron unos valores suficientes de enriquecimiento de la muestra final en la proporción de RNAs nacientes y fueron descartados. Otra aproximación que se testó fue el marcaje con Tio-UTP, su posterior acoplamiento reversible a moléculas de biotina y su purificación con columnas de estreptavidina. Esta estrategia se había testado en nuestro laboratorio anteriormente y se había descartado al revelarse un artefacto inesperado: el aparente marcaje inespecífico con residuos tiol de los RNAs maduros durante la reacción de *run-on*. Dada la utilidad mostrada por el tratamiento con RNasa para degradar los RNAs maduros antes del *run-on*, para esta tesis consideramos que al digerir con RNasa, este artefacto quedaría minimizado. A pesar de ello, y al igual que en los dos marcajes mencionados anteriormente, finalmente se descartó el uso del Tio-UTP al observar que los niveles de enriquecimiento finales no fueron suficientes como para descartar que la muestra no pudiera estar contaminada en gran medida con RNA maduro.

El último marcaje que se testó fue con Biotina-UTP. A diferencia de los marcajes descritos en el párrafo anterior, este procedimiento no permite una purificación selectiva eficiente, ya que la unión entre la biotina y la estreptavidina es tan fuerte que la elución desde las perlas de agarosa es muy poco eficiente (discutido con más detalle en la página final de la discusión), con lo que se pierde gran parte de la muestra inicial. Por ello, decidimos hibridar los RNAs nacientes biotinilados directamente. Los resultados muestran como, gracias a la congelación en nitrógeno de las células y a la digestión con RNasa previa al *run-on*, las señales de hibridación fueron globalmente mayores que en las aproximaciones anteriores con Biotina-UTP mencionadas. Además, se logró potenciar la señal global del chip realizando dos cambios adicionales en el protocolo: 1) separando los RNAs nacientes de los maduros por tamaño mediante el uso de columnas de purificación de RNAs pequeños y 2) duplicando el paso de tinción del chip.

Para descartar que el uso de marcaje interno de RNAs con biotina hiciera que hubiera un sesgo de intensidad de señal con la cantidad de uridinas de cada fragmento, se hizo un gráfico de correlación entre el porcentaje en G+C de cada sonda con su nivel de señal. Los resultados mostraron que no existe un sesgo de este tipo, probablemente debido a que el posible aumento de señal de las sondas más ricas en uridinas es contrarrestado por la peor capacidad de hibridación intrínseca que tienen estas secuencias ricas en A/T. En el

caso del GRO *clásico*, este sesgo se corrige multiplicando por el porcentaje relativo de uridinas de cada gen (García-Martínez *et al*, 2004).

Al visualizar los datos sobre el mapa genómico de la levadura se observó que las señales, aunque débiles en comparación a las de una hibridación de RNA maduro por el método convencional, eran apreciables y se localizaban mayoritariamente en los genes. La inspección visual detallada permitió encontrar diversos ejemplos demostrativos de la autenticidad de los datos de RNA naciente. Por ejemplo, los mostrados en la figura 11, en los que se aprecia señal antes y después de las regiones transcritas de los genes, y en las regiones intrónicas. Al analizar todos los genes en conjunto, surgieron otras evidencias que refuerzan la validez del método para capturar específicamente el nascentoma de la levadura. Entre ellas, la clara diferencia entre los perfiles globales de densidad de RNAPs alrededor del TSS y TTS con los perfiles derivados de análisis de RNAs maduros, o de CHIP de RNAPII (figuras 17 y 18). Otra evidencia surge de la comparación de las intensidades de señal globales antes y después del tratamiento con el inhibidor específico de la RNAPII α -amanitina (figura 34). En dicha figura se observa como la α -amanitina hace que las señales de los genes de RNAPII (sentido y antisentido) disminuyan considerablemente, mientras que no lo hacen así los genes de la RNAPIII, lo cual indica que las señales que estamos midiendo no son RNAs maduros existentes antes de la adición de la droga, sino RNAs que se sintetizan durante el *run-on*.

2. Ventajas del BioGRO frente al GRO *clásico* y a otras técnicas de RNA naciente.

Como se ha comentado anteriormente, a pesar de que el *run-on* es la técnica de elección para llevar a cabo estudios de medición de TRs nacientes, no está exenta de limitaciones. La actualización del método, que hemos denominado BioGRO, presenta las siguientes mejoras frente al GRO *clásico*:

- 1) **No se basa en marcajes radiactivos.** Este es un factor importante por tres razones fundamentales: la primera es que elimina el problema de tratar con

sustancias radiactivas, para cuya manipulación se necesita habitualmente tanto personal con la debida formación, como además la puesta a punto y el mantenimiento de instalaciones radiactivas adecuadas. La segunda es que el isótopo radiactivo utilizado en el GRO, ^{33}P , tiene una semivida relativamente corta (28 días), lo que obliga a un uso rápido del nucleótido comprado. La tercera es que posibilita la adopción de las plataformas genómicas de última generación, basadas en una miniaturización que las hace incompatibles con la captación de señales radiactivas.

2) **No sesgado por el diseño de las sondas.** La información se extrae de la totalidad del genoma gracias a la tecnología de microchips de embaldosado. De esta manera no está sujeta a limitaciones de diseño de las sondas como lo estaba el GRO, basado en sondas únicamente de ORFs (completas o parciales) de genes predichos. Ello permite obtener mucha más información no sesgada por la hipótesis previa.

3) **Genera mapas de transcripción a lo largo del genoma.** Ofrece la posibilidad de obtener datos promedio de nTR de genes al igual que el GRO, pero adicionalmente permite posicionar las RNAPs con mucha mayor resolución a lo largo del genoma.

4) **Es específico de hebra.** Al no basarse en la hibridación sobre sondas bicatenarias, ni en mecanismos de amplificación que supongan ningún paso de RT o PCR, se retiene la especificidad de hebra, muy útil para posibles estudios de caracterización de nuevas unidades transcripcionales no anotadas previamente, así como para un análisis de la transcripción no canónica (antisentido, críptica, etc.).

5) **Compatible con las plataformas de secuenciación masiva de última generación.** La naturaleza del método permite aislar fragmentos cortos de RNA, óptimos para construir librerías de secuenciación masiva. Con esta técnica se conseguiría aumentar tanto el rango dinámico como la resolución de los datos de *tiling arrays*.

A pesar de las mejoras, el BioGRO tiene la desventaja comparativa con el GRO *clásico* de rendir un nivel de señal de RNA naciente relativamente bajo, aspecto que debe mejorarse para poder extraer información de un número de genes parecido al del GRO.

En relación a las otras técnicas de medición de TR naciente (GRO-*seq* y derivados, y las basadas en fraccionamiento cromatínico), el BioGRO posee una ventaja importante: no requiere de ningún paso de purificación o amplificación del RNA naciente que pueda introducir ruido técnico en los datos, ya fuera este por contaminación de la muestra biológica con RNA maduro, o por sesgos introducidos al someter a la muestra a copiosos o amplificaciones enzimáticas.

3. Modelo de la autopista: ¿qué podemos aprender de los datos de BioGRO?

El proceso de transcripción de las RNAPs en general y de la RNAPII en particular puede modelizarse como el de la circulación de coches en una autopista, o mejor de trenes a lo largo de su vía de sentido único, dadas las evidentes similitudes que existen entre ambos (Dennis *et al.*, 2009; Tuller *et al.*, 2010). Aunque obviamente no es imprescindible esta comparación para explicar cómo transcriben las RNAPs, puede ayudar a la comprensión del proceso. Como en el caso de los coches en la autopista, la densidad de RNAPs en el gen depende de tres parámetros: 1) de la tasa de entradas (moléculas por unidad de tiempo), 2) de la velocidad promedio a la que se desplazan y 3) de la tasa de accidentes (*drop-off*, DO, en su caso). A su vez, la velocidad promedio resulta de la composición de velocidades variables, así como del número y duración de las paradas. El adelantamiento está prohibido (de ahí la analogía con el tren antes citada) pero son posibles los choques por alcance (Jin *et al.*, 2010). Una de las virtudes de la técnica BioGRO es su capacidad para describir la densidad de RNAP elongantes a lo largo del genoma lo cual, junto con otras medidas complementarias de RNAP totales (ChIP), velocidad promedio de RNAP (técnica del gen largo *YLR454W*, Mason y Struhl, 2005) y tasa de DO (gen largo y otras técnicas) posibilitan la descripción detallada del proceso de elongación y terminación que es función de las características del vehículo (complejo de elongación) pero también del asfaltado de la autopista (cromatina).

Tal y como se comenta en la introducción de esta tesis, la RNAPII es un complejo de 12 subunidades (holoenzima) que se ensambla en el promotor sobre secuencias de DNA que facilitan la formación del llamado *complejo de pre-iniciación* (PIC) junto con un numeroso grupo de factores adicionales, algunos específicos de gen y otros más o menos generales. El paso a complejo abierto e inicio de la polimerización a partir del sitio de inicio de la transcripción (TSS) se da a continuación. Sin embargo, las frecuentes paradas que se producen al inicio de la elongación con la generación de transcritos abortivos cortos hace considerar que la *entrada* real de las RNAPs en la autopista deba considerarse una vez pasado el punto en el que la elongación pasa a fase continua una vez fosforilada la Ser2 de la cola CTD. Podemos llamar tasa de inicio (TI) al número de moléculas de RNAPII que entran en fase de elongación productiva (>50 nucleótidos aguas abajo del TSS), y tasa de terminación (TT) al número de moléculas de RNAPII que llegan de forma productiva al sitio de poliadenilación dando lugar a un transcrito primario completo. Según esta definición, la TT equivale a la tasa de transcripción naciente (nTR) *real*.

Las RNAPs sufren paradas que no siempre pueden ser solventadas (los mecanismos de reactivación se detallan en la tabla 4). Por ello se puede decir que existe una tasa de accidentes o DO, que se puede definir como la diferencia entre TT y TI normalizada por unidad de longitud del transcrito (L):

$$DO = \frac{TT - TI}{L}$$

Según este razonamiento, la TI sería igual a la TT si no existiera DO de las RNAPs por el camino, en cuyo caso la DO sería 0. En otras circunstancias DO se mide en tanto por uno o tanto por cien de pérdida de RNAP elongantes por unidad de longitud (kb),

En los estudios de transcripción naciente, las TRs se suelen calcular haciendo dos asunciones o simplificaciones generales: velocidad constante y DO constante a lo largo de los genes. A pesar de ello, es muy probable que el DO no sea constante a lo largo de un gen, dado que los genes tienen diferentes secuencias de DNA y organizaciones cromatínicas que podrían influir en la frecuencia y/o duración de las paradas.

Por otro lado, las RNAPs se mueven a lo largo de los genes con una velocidad (V_i) variable. La variabilidad viene ocasionada por varias causas. La existencia de pausas y paradas del complejo de elongación provoca cambios bruscos de V_i en cada momento. Hay que recordar que los datos de nTR de BioGRO se basan en promedios de muchas células a la vez, con lo que con el fin de simplificar el proceso podemos considerar la velocidad como el promedio (V_p) de las V_i entre las múltiples copias del gen que se están estudiando. Si se considera que la V_p es constante para todos los genes y/o entre las condiciones que se comparan, la medición experimental de la densidad (D) de RNAPs que se obtiene a partir del BioGRO es una buena medida de la nTR. Este tipo de consideración ha sido ampliamente utilizada tanto a escala individual (Hirayoshi y Lis, 1999) como a escala genómica (García-Martínez *et al.*, 2004; Mayer *et al.*, 2010).

A falta de datos genómicos fiables de V_p que sirvieran para refinar los cálculos de nTR, los datos aparecidos recientemente (Miller *et al.*, 2012) que muestran un valor de DO global en levadura distinto de cero, nos sirvieron para corregir nuestros datos de nTR *aparentes* (tanto de GRO *clásico* como de BioGRO) y acercarlos a los *reales*. El efecto de la corrección se hizo visible de dos maneras: 1) al observar como la nTR se hace progresivamente menor conforme aumenta la longitud de los genes (figura 15), y 2) al ver que los datos de BioGRO se asemejan más a los de SR (calculados por DTA, figura 15). El primer punto refleja que el tratamiento de los datos corrige la sobre-estimación de las señales de los genes largos que se produce por considerar que toda RNAP elongante detectada será capaz de producir un transcrito completo, sin tener en cuenta que el DO existente predice que algunas de ellas caerán del DNA sin acabar la transcripción. Lo cual será más probable cuanto más largo sea el gen transcrito. El segundo punto refleja que, una vez corregida la transcripción naciente, ésta se parece más a la tasa de síntesis real de mRNAs maduros (SR).

4. Características generales del gen tipo de RNAPII. Subtipos.

El análisis del *Gen Modelo* aplicado a los datos de BioGRO nos permitió dibujar un perfil global de densidad de RNAPs activas a lo largo de los genes. Dado que la presencia de

RNAPs activas es esperable más allá de las zonas de inicio y final de la zona codificante que se utilizan normalmente para los estudios transcriptómicos, en esta tesis usamos anotaciones de inicio y final de transcripción (sitios TSS y TTS). El perfil global de RNAPs activas revela la existencia de una meseta más o menos estable de señal 3 veces por encima del ruido, con una ligera acumulación de señal hacia el extremo 3'. Dicha meseta está flanqueada a ambos lados por dos mínimos de señal coincidentes con las regiones 5' promotora y 3' terminadora de los genes (figura 16, A). Gracias a que los microchips utilizados son informativos para las dos hebras del DNA del genoma, y que el método de BioGRO es específico de hebra, fuimos capaces de dibujar el perfil de señal en antisentido para todos genes (figura 16, A). De él se puede extraer la conclusión de que este tipo de transcripción es minoritaria respecto a la canónica en *S. cerevisiae*, aunque la baja señal general obtenida con el método dificulta hacer más valoraciones a este respecto.

El perfil de los genes con intrón reveló la existencia de una acumulación de señal de RNAPs activas hacia la zona 3' de los genes, compatible con las recientemente descubiertas dinámicas de *splicing* co-transcripcional en levadura. Nuestros datos se suman a las evidencias de que las RNAPs se pausan o disminuyen su velocidad al principio del segundo exón para permitir a la maquinaria de corte y empalme procesar el intrón antes de reanudar la transcripción (Carrillo-Oesterreich *et al.*, 2010; Alexander *et al.*, 2010; Braberg *et al.*, 2013).

El análisis del metagen de extremos 5' y 3' permitió un estudio más detallado de las zonas de inicio y final de la transcripción. Dicho análisis consistió en alinear todos los genes por el TSS o el TTS y obtener el perfil promedio de densidad de RNAPs activas en una región fija de 500 pb alrededor de dichos puntos. En la figura 17 (A y B) se comparan los perfiles promedio de BioGRO con los de un RNA total (maduro), viéndose claramente la diferencia de perfil, tanto en el TSS como en el TTS, y poniéndose de manifiesto que el perfil de RNAPs activas tiene unas características peculiares que ya se intuían en el análisis del *Gen Modelo*. A grandes rasgos, ambos métodos reflejan que los puntos de menor señal corresponden a las regiones justo antes y después del TSS y TTS, respectivamente, pero es en el perfil del metagen donde se aprecia como la región 5' del gen no posee un perfil plano, sino que la señal varía de intensidad con varios picos en la zona inicial del transcrito. Asimismo, la región del final del gen presenta una doble acumulación de señal

antes del TTS que no se apreciaba en el perfil del *Gen Modelo*, y que podría estar reflejando las regiones donde las RNAPs disminuyen su velocidad a consecuencia de la actuación los mecanismos de corte y poliadenilación.

El patrón oscilante de la señal de BioGRO en la zona 5', con un espaciado regular de aproximadamente 165 nt, nos hizo pensar en una posible vinculación con la disposición de los nucleosomas sobre el DNA. Tal razonamiento vino motivado en parte por la existencia de diversos estudios en los que se habla de las dinámicas de transcripción de las RNAPs a través de la cromatina, en las que el principal rasgo característico es que los nucleosomas presentan un obstáculo para las RNAPs en transcripción, a las cuales inducen a pararse y a hacer *backtracking* (Hodges *et al.*, 2009, Jin *et al.*, 2010, Chung *et al.*, 2010, Churchmann y Weissman, 2011). Al superponer el perfil promedio de nucleosomas sobre la región 5' (figura 17, C) se observa claramente que los mínimos de señal de BioGRO en esa zona coinciden con máximos de posicionamiento de nucleosomas, indicativo de que las RNAPs tienen mayor densidad (por lo tanto que avanzan más lento) en las zonas entre nucleosomas. A este perfil característico lo consideramos por tanto un perfil *antinucleosomal*. Dicho perfil no se observó tan claramente en la zona 3' de los genes, debido probablemente a que en esta zona los nucleosomas están menos posicionados que en 5' (Lee *et al.*, 2007, Jiang y Pugh, 2009, Weiner *et al.*, 2010, Brogaard *et al.*, 2012).

Una de las técnicas de medida de la transcripción naciente más relevantes y más utilizadas en la actualidad en los estudios de genómica funcional es la técnica de ChIP, que como se ha comentado mide la presencia de RNAPs en contacto con el DNA (Aparicio *et al.*, 2004; Waldminghaus y Skarstad, 2010). A diferencia del BioGRO, en el que la señal proviene únicamente de RNAPs competentes para la elongación, la naturaleza de la técnica de ChIP hace que la señal (que en los datos de Lola de Miguel y Sebastián Chávez que hemos utilizado en este trabajo se obtuvo mediante IP de la subunidad Rpb3 de la RNAPII) no pueda atribuirse a un estado transcripcional determinado, informando únicamente de la presencia de una RNAPII en esa región. La comparación directa por superposición de perfiles (en figuras 18, A y B) reveló que:

- 1) el patrón *antinucleosomal* en 5' es específico de las RNAPs activas
- 2) mientras que el patrón de ChIP en 3' del gen refleja la dinámica de formación del extremo 3' y la terminación, lo que se conoce como *transición 3' en dos pasos* (Mayer *et al.*, 2010), el perfil de BioGRO presenta características distintivas.

La relación o el *ratio* entre las señales permitió destacar los perfiles específicos de RNAPs elongantes (figuras 18, C y D) o paradas (figuras 18, E y F). Como era de esperar, el perfil de RNAPs elongantes sirvió para acentuar las características del perfil de BioGRO, viéndose ahora más nítidamente los picos de acumulación antes de cada nucleosoma en la región 5', y un gran pico de acumulación justo antes del sitio de pA. El *ratio* inverso nos informó sobre la presencia de RNAPs paradas, que se revelaron como era de esperar acumuladas justo sobre las zonas donde se posicionan los centros de los nucleosomas en 5' (figura 18, E). Además, este perfil aporta evidencias adicionales a favor de que no existe una acumulación de RNAPs paradas en la zona justo después del inicio de la transcripción, un mecanismo generalmente aceptado en metazoos, pero que también había sido descrito en *S. cerevisiae* (McKinlay *et al.*, 2011).

Finalmente, el análisis del perfil en la región 3' reveló que las RNAPs no hacen *backtracking* durante la poliadenilación (coincide con un mínimo de RNAPs paradas), pero sí justo después, durante la terminación (donde vuelve a aparecer un pico de RNAPs paradas).

Una vez desvelado el patrón promedio de transcripción naciente para el conjunto de genes, quisimos ver si encontrábamos genes con patrones distintos al promedio, y si habían grupos con perfiles característicos compartidos en 5'. La primera aproximación llevada a cabo (la búsqueda no dirigida de subgrupos) no resultó eficiente, probablemente debido a la baja señal general. No obstante, sí que nos sirvió para confirmar lo visto tanto en genes individuales como en los perfiles del metagen: que al hacer el *ratio* entre las señales de BioGRO y RNA total y ordenar los genes según su tamaño, existía más señal de BioGRO que de RNA total en las zonas antes del TSS y después del TTS (figura 21). La extensión en 5' parece estar reflejando la existencia de transcritos crípticos en el mismo sentido de la transcripción pero aguas arriba del TSS canónico, y la del 3' apunta a la existencia de sitios de pA alternativos al mayoritario.

Además, nos llamó la atención que la extensión de la señal parecía ser más larga en la zona más allá del pA de los genes más cortos, cosa que se confirmó al hacer un estudio más detallado de perfiles del metagen de grupos de genes cortos. Esta extensión por 3' preferiblemente en los genes cortos cuadra con la hipótesis del *gene looping* (O'Sullivan *et al.*, 2004; Ansari y Hampsey, 2005; Grosso *et al.*, 2012, Al Husini *et al.*, 2013), que postula que los determinados grupos de genes quedan con las regiones promotora y terminadora yuxtapuestas a modo de lazo tras una primera ronda de transcripción, para permitir la iniciación rápida de una segunda ronda por la misma RNAP responsable de la ronda pionera. La hipótesis propone que para la formación del lazo es necesario que la conexión entre el terminador y el promotor no se dé justo entre el TSS y el pA, sino que se unan tomando una extensión concreta de la zona después del pA. Por ello, la extensión más allá del sitio de pA que se observa en BioGRO sería el resultado de la presencia de RNAPs activas transcribiendo la porción del terminador que sirve de puente entre la terminación y el reinicio.

La segunda aproximación, dirigida a extraer los perfiles de grupos concretos de genes con características comunes o cuya regulación se sabe compartida, mostró que el perfil general puede dividirse en varios sub-perfiles. Así, al dividir los genes arbitrariamente en tres categorías según su nivel de transcripción, se vio que el patrón general es el reflejo fundamentalmente del perfil de los genes de alta transcripción, y que los genes de media o baja transcripción presentan un perfil ruidoso que aporta poco al perfil general (figura 20, A). De esta manera, los genes RP (como ejemplo de genes co-regulados y de alta transcripción) se revelaron como el perfil paradigmático de densidad de RNAPs activas sobre los transcritos, tanto en 5' como en 3'. En él se recogen las características principales anteriormente descritas: mínimos de señal coincidiendo con los NFRs, patrón claramente *antinucleosomal*, y doble acumulación antes del sitio pA. Por su parte, la distinción entre genes con promotores con o sin caja TATA canónica reveló que cada uno tiene un perfil de RNAPs activas distinto (figura 20, C), indicativo de la existencia de programas de expresión génica particulares dependiendo del tipo de promotor. El perfil de los genes TATA, distinto al del conjunto de genes y con un mayor nivel de densidad de RNAPs elongantes, llama la atención ya que no se había descrito en la literatura que los genes TATA tuvieran un nivel de transcripción mayor al resto de genes. Revisando sin embargo otros conjuntos de datos publicados nos dimos cuenta de que, efectivamente,

los genes TATA se transcriben más que los TATA-*like* y tienen también un mayor nivel de mRNA (tabla 3). Este resultado es, para nuestro conocimiento, novedoso, y supone que el grupo de genes con caja TATA, que incluye muchos genes de respuesta al estrés que tienen bajo nivel de transcripción en las condiciones estudiadas (crecimiento exponencial en YPD), se compone también de otros genes de alta transcripción en condiciones óptimas de crecimiento.

5. Estudio del nascentoma de los mutantes *rpb4Δ* y *xrn1Δ*.

Una vez caracterizado el nascentoma de una cepa salvaje y con el fin de conocer sus posibilidades, decidimos aplicar la técnica a cepas deficientes en factores relacionados directa o indirectamente con la transcripción: los mutantes *rpb4Δ* y *xrn1Δ*. La primera cepa carece de una subunidad pequeña no esencial de la RNAPII, la subunidad Rpb4. Esta se asocia *in vivo* con la otra subunidad pequeña, Rpb7, y juntas componen un sub-módulo funcional que interviene por una parte activando la transcripción y por otra estimulando la degradación de los mensajeros maduros en el citoplasma. Se considera por tanto como un ejemplo claro de *coordinador de mRNA* (Choder *et al.*, 2010, Shalem *et al.*, 2011). La otra cepa que se caracterizó fue la cepa *xrn1Δ*, deficiente en el principal factor de degradación de los RNAs celulares (Garneau *et al.*, 2007; Parker, 2012; Pérez-Ortín *et al.*, 2013). El interés por esta proteína deriva, sin embargo, en que también se le ha atribuido recientemente un papel como estimulador de la transcripción. Esto hace que se la considere como una *sintegradasa* (Haimovich *et al.*, 2013), proteína con doble función en síntesis y degradación de mRNAs.

La representación de los datos de BioGRO del mutante *rpb4Δ* frente a los del WT nos permitió detectar una caída de la nTR relativa con la longitud de los genes (figura 23, B). Estos resultados sugieren la existencia de defectos en elongación por parte de la cepa carente de Rpb4, y confirman hallazgos previos obtenidos en nuestro laboratorio mediante la técnica del GRO radiactivo (Daniel Medina, comunicación personal). El análisis del metagen en los extremos 5' y 3' mostró un perfil diferente al de la cepa salvaje, tanto por el nivel global de intensidad de señal, como por la pérdida de señal

progresiva que se observa después del TSS (figura 23, C). Esta pérdida de señal podría interpretarse, no como un defecto de elongación, sino como un aumento de la velocidad de elongación de las RNAPs que hace que sean capturadas en esa región con menos probabilidad que las RNAPs más *lentas* del WT. Sin embargo, y a la luz de los resultados del análisis mencionado más arriba, este patrón cuadra más con la existencia de un defecto de elongación. Tal defecto hace que las RNAPs tengan dificultades para progresar y en parte se suelten del sustrato (DO) en ausencia del factor Rpb4. Además, estudios recientes del laboratorio de S. Chávez (comunicación personal) indican que el mutante Rpb4 presenta una velocidad de elongación promedio sólo un poco inferior a la del WT, así como un aumento significativo del DO en el gen largo *YLR454W*, lo cual apoya nuestra interpretación.

En cuanto a la cepa carente de Xrn1, cabe destacar que se obtuvieron unos marcajes muy pobres, indicativos de que la transcripción en estas células está muy afectada por la ausencia del factor. Este resultado coincide con los datos obtenidos con GRO radiactivo (Haimovich *et al.*, 2013). Este bajo nivel de señal obtenido hace que la capacidad del método BioGRO se vea limitada, pudiéndose extraer pocas conclusiones claras al respecto. Cabe destacar en este sentido que el perfil del metagen refleja la ausencia de patrón *antinucleosomal*, probablemente debida a la baja señal general obtenida, que hace que el perfil sea más ruidoso y menos claro.

Por otro lado, los hallazgos recientes en la literatura sobre el papel de Xrn1 en el control de la transcripción antisentido (van Dijk *et al.*, 2011) motivaron que analizáramos ambos conjuntos de datos de BioGRO. A este respecto cabe destacar el aparente aumento en el nivel de transcripción antisentido global desde el NFR 3' hacia dentro de los genes (figura 24 B) y el aumento del nivel de transcripción naciente de los transcritos XUTs. Estos resultados, en conjunto, indican que la ausencia de Xrn1 afecta a la ocupación de RNAPII sobre los genes, a escala genómica. Este efecto parece ser doble: por una parte disminuye la transcripción canónica (en sentido), y por otra parte (o quizá como consecuencia de esto) aumenta la transcripción antisentido y críptica. De nuevo hay que dejar claro que estos resultados hay que tomarlos como preliminares, dado que el bajo nivel de señal podría estar enmascarando otras características interesantes que pudieran alterar las conclusiones.

6. RNAPII y nucleosomas, interdependencia.

Tal y como se discute en el apartado de caracterización general del gen de la RNAPII, las discontinuidades de señal apreciables en el perfil promedio de RNAPs activas sobre la región inicial del gen parecen estar reflejando la manera que tiene la RNAPII de transcribir moldes cromatínicos, y como los nucleosomas imponen barreras para el avance de las RNAPs en transcripción. Esta hipótesis ya había sido debatida en la literatura, tanto con estudios *in vitro* como *in vivo* en levadura, llegándose a la conclusión de que efectivamente la tasa de elongación de las RNAPs se ve afectada por la presencia de los nucleosomas que cubren las regiones a transcribir (Hodges *et al.*, 2009; Churchman y Weissman, 2011). El uso del método BioGRO permite testar esta hipótesis desde una perspectiva distinta pero complementaria. Para esta tesis se eligieron diferentes aproximaciones para testar la hipótesis de si la densidad de RNAPs activas se ve afectada por los nucleosomas. Por una parte, se buscó esta interdependencia comprobando si genes con distintos patrones de posicionamiento nucleosomal tenían perfiles distintos de BioGRO, y, por otra parte, mediante el estudio del nascentoma de mutantes relacionados con la transcripción a través de moldes cromatínicos.

Los resultados de la primera aproximación revelaron la existencia de grupos de genes con diferentes patrones compartidos de posicionamiento de nucleosomas (figura 25), y pusieron en evidencia como las diferentes configuraciones nucleosomales condicionan el patrón de RNAPs activas sobre esos genes (figura 26).

Para abordar la segunda aproximación se eligieron dos cepas distintas: un mutante carente del remodelador de cromatina *lsw2*, y otro carente de *TFIIS/Dst1*. De los análisis del fenotipo general del mutante *lsw2* cabe destacar la existencia de un posible defecto de elongación progresivo conforme aumenta la longitud de los genes, que aparece al hacer la relación entre la transcripción del mutante y el WT y representarla frente a la longitud de los genes (figura 27, B). Sin embargo, no se observaron diferencias notables entre los perfiles del metagen 5' y 3' entre mutante y WT, más allá de una ligera bajada de la intensidad general de la señal, y un ligero desplazamiento del patrón del mutante hacia 3' (figura 27, C). Estos últimos resultados entraban dentro de lo esperable, ya que se

ha descrito que el factor *Isw2* ejerce su función sobre un grupo de genes que no supera el 12% del total. Debido a que nuestros análisis se basan en dibujar un perfil promedio de todos los genes, esta proporción tan baja de genes afectados por la ausencia del factor queda probablemente enmascarada por el 88% restante de genes cuyo patrón de nucleosomas no depende de *Isw2*. Para refinar la búsqueda quisimos ver si existían diferencias entre los perfiles del metagen 5' de mutante y WT en grupos más concretos de genes. Para ello nos basamos en listas de genes obtenidas de estudios publicados: por una parte, una lista con aquellos genes que habían sido definidos como *dependientes* de *Isw2* (Whitehouse *et al.*, 2007), y por otra parte, una lista de genes a los que se *une* *Isw2* (Yen *et al.*, 2012). Como se muestra en la figura 28, el solapamiento entre estas dos listas publicadas es sorprendentemente bajo, lo que parece reflejar que la detección de la unión de *Isw2* a un gen no necesariamente refleja un efecto posicionador importante por su parte y, por otro lado, que un cierto número de genes pueden tener nucleosomas posicionados de forma indirecta por *Isw2*. Las únicas diferencias entre mutante y WT que encontramos con nuestros datos de BioGRO aparecieron al dibujar el perfil de los genes *unidos* por *Isw2* (figura 28, B). A pesar de que esta coincidencia sería suficiente para permitirnos concluir que la ausencia de *Isw2* tiene un impacto en el patrón de RNAPs activas indirectamente a través de la desorganización de los nucleosomas de esa región, decidimos no dar por concluido el estudio y confirmar los resultados mediante otra aproximación. Dicha aproximación consistió en tomar los datos publicados de posicionamiento de nucleosomas de las cepas mutante *isw2* y WT y someterlos a un reanálisis global con nuestros procedimientos. La búsqueda de grupos de genes cuya diferencia de posicionamiento de nucleosomas fuera más acusada entre mutante y WT nos permitió encontrar un grupo de 177 genes cuyo desplazamiento nucleosomal hacia 3' en ausencia de *Isw2* se reflejaba con claridad también en los perfiles de BioGRO (figura 30). Así pues, tanto los genes que son dianas preferentes de *Isw2* (Yen *et al.*, 2012), como el grupo de genes con mayor dependencia de *Isw2* en su posicionamiento nucleosomal (según nuestro estudio) sufren un desplazamiento equivalente de patrón de picos de RNAP elongantes. Esto demuestra que esos picos de RNAP son dependientes del posicionamiento de, al menos, los primeros 3 nucleosomas de la región transcrita. En resumen, la acumulación de RNAPs elongantes entre nucleosomas y los mínimos coincidentes con los nucleosomas posicionados sugieren un mecanismo de elongación en

el que las RNAPs reducen su velocidad al aproximarse a los nucleosomas, posiblemente por el obstáculo topológico que representan. Una vez han eliminado el obstáculo, bien por los mecanismo de \emptyset -looping (Kulaeva *et al.*, 2009; Luse y Studitsky, 2011), bien por la eliminación parcial o completa de histonas (Kireeva *et al.*, 2002, Studitsky *et al.*, 1997), se produce un aumento de la velocidad hasta encontrarse con el siguiente obstáculo.

Otra manera, complementaria a la anterior, de evaluar la interdependencia entre nucleosomas y RNAPs fue el análisis de la cepa carente de TFIS/Dst1. Este factor es el encargado de ayudar a reactivar las RNAPs que entran en estado de *backtrack*, situación que ocurre con frecuencia cuando una RNAP *choca* con un nucleosoma. La hipótesis que quisimos testar es si la ausencia del factor se reflejaría en una desaparición del patrón *antinucleosomal*, como consecuencia de que las RNAPs paradas no fueran capaces de reiniciar (dando así otro pico posterior de señal de BioGRO). Los resultados (figura 32, C) mostraron que el patrón nucleosomal se aplana y se hace menos evidente, pero no lo suficiente como para afirmar con seguridad que existe un efecto global. Lo que sí se encontró fue un defecto de elongación que se asemeja al encontrado en el mutante Rpb4 (figura 23, B). Siguiendo el mismo razonamiento que para el mutante *lsw2*, decidimos refinar la búsqueda del fenotipo, analizando para ello los perfiles de grupos de genes en los que se había publicado que el factor Dst1 tenía un papel especialmente relevante. Entre ellos encontramos los genes RPs y los genes TATA (Gomez-Herreros *et al.*, 2012). En ambos casos sí se encontraron diferencias claras entre los perfiles (figura 33), indicativas de que el mutante posee defectos de elongación, sobre todo en genes largos o de alta transcripción.

7. Nascentoma de RNAPIII.

En esta tesis se muestran los datos del primer estudio global en un eucariota de la transcripción naciente de la RNAPIII. Una primera aproximación a su caracterización fue llevada a cabo mediante la realización de un experimento de inhibición selectiva de la transcripción usando la micotoxina α -amanitina. El estudio comparativo de las señales de transcripción naciente tras una reacción de *run-on* en presencia o ausencia de la toxina

nos permitió separar la contribución de las diferentes RNAPs a la señal obtenida. Según se muestra en la figura 34, la señal de los genes de RNAPII (sentido y antisentido) cae tras el tratamiento, mientras que la de los genes de la RNAPIII se mantienen en niveles similares al control, con la excepción del gen *snR6*, lo cual podría estar indicando que o bien este gen está erróneamente asignado como transcrito por la RNAPIII, o bien solapa con una región de RNAPII que sí es inhibible por α -amanitina. Como ya se ha mencionado anteriormente, estos resultados indican que la α -amanitina no es capaz de bloquear completamente la transcripción de la RNAPII (Rudd y Lusse, 1996) y que el BioGRO mide inequívocamente transcripción nascente. Por otra parte, permiten afirmar que el método es capaz de caracterizar la transcripción de la RNAPIII, ejemplificada por los altos niveles de transcripción que exhiben los genes de tRNAs. De los mismos datos se extrae la conclusión de que la transcripción antisentido en los genes de tRNAs es muy baja, indistinguible del ruido técnico del método. Esto cuadraría con lo que se ha observado en otros genes de alta transcripción, para los que se ha descrito que su transcripción estaría condicionada o favorecida en un único sentido (Tan-Wong *et al.*, 2012).

El análisis del metagen desde el extremo 5' (figura 35) dibuja un perfil de RNAPIII activa sobre los tRNAs muy característico, en forma de fuerte pico de señal 50 veces mayor que el promedio para los genes de RNAPII, que desciende desde aproximadamente la mitad del transcrito modelo. Este descenso puede ser atribuible a un aumento de la velocidad de las RNAPs hacia el final de los genes. Asimismo, el perfil de densidad no se ve prácticamente afectado por el tratamiento con α -amanitina, tal y como se apreciaba también en los datos de la figura 34.

Algunos tRNAs poseen intrones en su secuencia, pero a diferencia de los intrones de los genes de RNAPII, los de tRNAs no se procesan co-transcripcionalmente. Esta diferencia fundamental hace que no se esperen diferencias *a priori* entre los perfiles de densidad de RNAPs activas en los tRNAs con y sin intrones. Sin embargo resultó que sorprendentemente sí existían diferencias entre ambos perfiles (figura 36). Los genes con intrón se caracterizan por tener dos picos de acumulación de RNAPs activas dentro de la región transcrita: el primer pico coincide con el de todos los genes, y el segundo coincide con la región de unión intrón-exón 3'. Dado que no parece lógico pensar que ello sea reflejo de una ralentización asociada a mecanismos de maduración del transcrito,

decidimos hacer un estudio de la secuencia del DNA de los tRNAs y vimos que la región intrónica coincide con una zona de alto contenido en AT, asociada a una baja estabilidad del complejo de elongación (Kraeva *et al.*, 2007). Esta baja estabilidad del complejo estaría afectando a la velocidad de la RNAPIII, haciendo que apareciera un segundo pico de acumulación hacia 3', que luego vuelve a desaparecer conforme las RNAPs acaban de transcribir la zona del intrón y ganan velocidad.

La medición de la transcripción nascente de los tRNAs posibilitó obtener una idea más amplia de cómo se regula la expresión de estos genes. La comparación de la señal promedio de cada familia de tRNAs tanto con su número de copias, como con la concentración a la que están presentes en el citoplasma (figura 37) mostró correlaciones positivas. Por otra parte, al analizar si las familias *favoritas* (seleccionadas positivamente por la selección traduccional) tenían una transcripción nascente promedio más elevada que las *no favoritas* se encontró que no había diferencias significativas. Estos datos tomados en conjunto indican que la nTR es aproximadamente constante por copia de isógen de tRNA, y que el hecho de que una especie de tRNA esté más o menos disponible para la maquinaria de traducción es consecuencia de una transcripción incrementada solo por el hecho de poseer un mayor número de copias dispersas por el genoma.

También se analizó el posible papel de los factores Dst1 e Isw2 en la expresión de los genes de tRNAs. A ambos factores se les ha asignado recientemente un función importante en la regulación de este tipo de genes (Ghavi-Helm, 2008; Kumar y Bhargava, 2013): Dst1 como GTF, e Isw2 como remodelador de cromatina general de los tRNAs. El análisis de nuestros datos de nTR en ambos mutantes simples mostró un fenotipo de alteración general en los patrones de TR de los tRNAs, mayor incluso del que se había descrito en la literatura (figura 39). De entre los genes alterados, encontramos dos grupos de tRNAs cuya expresión se vio en general reprimida en los mutantes, y después un par de genes sobre-expresados en el mutante Dst1 y otro par distinto en el mutante Isw2. Estos resultados podrían deberse en parte simplemente al hecho de que los mutantes crecieran peor que el WT, lo cual afecta tanto a la actividad de la RNAPI como de la RNAPIII (Hall *et al.*, 2006). No obstante, no parece ser el caso de estas cepas, ya que su tasa de crecimiento es muy similar a la cepa WT. Por tanto, estos datos en conjunto apoyan la idea de que tanto Dst1 como Isw2 tienen un papel importante en la regulación

de la expresión de los tRNAs, que podría indicar la existencia de una mayor relación de la que se pensaba entre las maquinarias de las RNAPs II y III.

En este sentido, realizamos un estudio global de la relación entre ambas maquinarias, midiendo para ello la nTR promedio de los genes de tRNAs con un transposón insertado en su vecindad, y comparándola con los que no lo tuvieran. Los datos mostraron que, a pesar de la proximidad de las unidades transcripcionales, las nTR promedio entre los dos grupos no eran significativamente distintas (figura 40). Dado que se sabe que los tRNAs establecen un estado de cromatina represiva en sus zonas flanqueantes que impide la transcripción de genes próximos de RNAPII excepto de los de transposones, que se expresan a niveles bajos (mecanismo de silenciamiento tgm, ver Introducción), lo que indican los resultados es que este efecto de silenciamiento condiciona a la maquinaria de RNAPII, poco presente en los genes de retrotransposones, impidiendo a su vez posiblemente que compita con la de RNAPIII encargada de transcribir fuertemente los tRNAs.

Una última consideración sobre los tRNAs que no tiene una relación directa con los datos de BioGRO, pero a la que no se había hecho referencia antes en la literatura, es la dependencia de la presencia de un transposón insertado preferentemente en la vecindad de los genes de las familias con mayor número de copias. Según esta hipótesis, esta presencia preferente sería el resultado de un mecanismo evolutivo por el cual los tRNAs se dispersarían por el genoma con la ayuda de un mecanismo mediado por la movilización de sus retrotransposones asociados. Los datos muestran sin embargo que la hipótesis no parece ser cierta, y que los tRNAs *favoritos* no contienen retrotransposones en su vecindad con mayor frecuencia que los *no favoritos* (figura 41).

8. Posibilidades de mejora de la técnica y aplicación a otras levaduras.

Tal y como se ha discutido hasta el momento, el BioGRO es capaz de caracterizar el nascentoma de *S. cerevisiae* de manera efectiva y no sesgada. A pesar de ello, en el punto actual de desarrollo la técnica adolece de dos limitaciones principales: 1) la baja sensibilidad global, lo cual hace que sólo se tengan datos robustos de una población de

alrededor de 1000 genes de la RNAPII, y 2) que no está adaptada todavía a las plataformas de secuenciación masiva que se utilizan en la actualidad. Ambas limitaciones están siendo abordadas en nuestro laboratorio actualmente, con resultados prometedores. Por una parte se están probando estrategias de marcaje con biotina, purificación con esferas de estreptavidina y nuevos métodos para mejorar la elución, y por otra parte se ha desarrollado un protocolo de construcción de una librería de secuenciación masiva con la plataforma Illumina a partir de los RNAs de BioGRO. El éxito de estas aproximaciones permitirá ir más allá en la caracterización de los genes de media y baja transcripción (incluyendo los antisentido y demás transcripción no canónica), a una mayor resolución (en el caso de la ultrasecuenciación).

Otra de las posibilidades del método es su aplicación a otros organismos, además de a *S. cerevisiae*. En nuestro laboratorio se ha testado el BioGRO en la levadura patógena de humanos *Candida albicans*, con resultados positivos (Ana Miguel, comunicación personal). La efectividad del BioGRO en *C. albicans* abre la puerta al uso de la técnica en otros organismos, como *Drosophila*, ratón o humanos.

CONCLUSIONES

1. Desarrollo del método BioGRO.

1.1 Se ha puesto a punto un método para caracterizar la transcripción nascente a escala genómica en la levadura *S. cerevisiae*. La técnica permite posicionar RNAPs elongantes de manera no sesgada a lo largo del genoma, en ambas hebras, a una resolución mayor que la obtenida con metodologías anteriores.

1.2 El método es útil para la estimación de tasas de transcripción nascente de manera fiable en los genes de alta transcripción. Los datos obtenidos por BioGRO proporcionan una visión nueva sobre los mecanismos de regulación de la expresión génica en levadura, complementaria a la que se obtiene mediante las otras técnicas de RNA nascente existentes.

1.3 El gen *tipo* transcrito por la RNAPII posee un perfil de RNAPs elongantes variable a lo largo de la región transcrita, cuyas características generales son: mínimos de señal coincidentes con las regiones NFR promotora y terminadora, oscilaciones con patrón de espaciado regular al inicio del gen, y doble acumulación de RNAPs activas hacia la región 3'. Dentro de este patrón general existen grupos concretos de genes con perfiles diferentes, como los genes con intrones, genes cortos, o genes TATA.

2. Aplicación del método a diferentes cepas de levadura.

2.1 El mutante *rpb4Δ* posee un patrón característico, indicativo de un claro defecto de elongación.

2.2 La cepa carente del gen *Xrn1* presenta severos defectos en transcripción. Además, presenta niveles aumentados de transcripción antisentido. Estos datos refuerzan el papel de *Xrn1* como efector positivo de la elongación y, por lo tanto, como coordinador global de la homeostasis del mRNA.

3. Relación entre RNAPs activas y cromatina.

3.1 La comparación de mapas genómicos de nucleosomas y RNAPs elongantes en la cepa salvaje permiten concluir que existe una compleja relación de interdependencia entre ambos complejos macromoleculares, que determina su

posicionamiento sobre el DNA. Por una parte, los nucleosomas presentan un serio obstáculo para las RNAPs en transcripción activa, y por otra parte, el paso de las RNAPs determina tanto el posicionamiento como la tasa de recambio de los nucleosomas.

3.2 La ausencia del factor remodelador de cromatina Isw2 resulta en un desplazamiento hacia 3' de los primeros nucleosomas de un grupo específico de genes, los cuales presentan también el mismo desplazamiento del patrón de densidad de RNAPs elongantes. Esto demuestra que el perfil oscilante observado en la región 5' del gen tipo se debe a acumulaciones locales de RNAPs antes de cada nucleosoma.

3.3 El factor de transcripción TFIIS es necesario para la correcta transcripción de moldes cromatínicos, especialmente las regiones 5' de los genes largos y los genes de alta transcripción.

4. Caracterización del nascentoma de la RNAPIII.

4.1 El método BioGRO proporciona el primer conjunto de datos de tasas de transcripción nacientes de la RNAPIII a escala global en un eucariota.

4.2 Los genes de tRNAs se caracterizan por una elevada tasa de transcripción naciente canónica y una baja tasa de transcripción naciente antisentido. Su perfil promedio de densidad de RNAPs elongantes se sitúa en medio de una región desprovista de nucleosomas.

4.3 La tasa de transcripción de RNAPIII es aproximadamente constante por isogen de tRNA, tanto para familias génicas *preferidas* como *no preferidas* por la maquinaria traduccional. Esto, unido al hecho de que los tRNAs *preferidos* poseen un mayor número de copias en el genoma, permite concluir que, en la levadura *S. cerevisiae*, la evolución ha optimizado la traducción mediante la duplicación génica de los tRNAs *preferidos*.

4.4 La ausencia de Isw2 y de TFIIS tiene como consecuencia un patrón de expresión de tRNAs alterado, lo cual indica que ambos factores tienen un papel importante en la regulación de la transcripción de este tipo de genes.

4.5 Los elementos transponibles no interfieren de manera importante sobre la transcripción de los tRNAs.

LISTADO DE TABLAS Y FIGURAS

Introducción

Figura i1	Estructura del holoenzima de la RNAPII.	5
Figura i2	Modelo estructural del complejo de promotor mínimo y co-activadores.	10
Figura i3	Señales de procesamiento del extremo 5' del pre-mRNA y complejos de C/P.	19
Figura i4	Esquema del proceso de la expresión génica que muestra las diferentes técnicas de medición de nTR y SR.	34
Figura i5	Modelo estructural del holoenzima de RNAPIII.	37
Tabla i1	Relación de subunidades de la RNAPIII y sus homologías con la RNAPII.	38
Figura i6	Tipos de promotores de RNAPIII en levadura.	40

Materiales y Métodos

Tabla 1	Cepas de <i>S. cerevisiae</i> usadas en este trabajo.	57
Tabla 2	Oligonucleótidos utilizados en este trabajo.	64

Resultados

Figura A	Resultados preliminares de un GRO no radiactivo.	88
Figura 1	Perfil electroforético de RNAs de levadura antes y después de tratamiento con RNasa A.	90
Figura 2	Incorporación de nucleótidos radiactivos mediante GRO en diferentes condiciones.	92
Figura 3	Aumento del porcentaje de incorporación radiactiva tras congelación.	93
Figura 4	Esquema de las distintas aproximaciones empleadas en este trabajo para el marcaje y enriquecimiento de RNAs nacientes.	94
Figura 5	Selectividad de la biotilación de RNAs tiolados.	96
Figura 6	Esquema del método BioGRO.	100
Figura 7	Análisis del tamaño de elongación de run-on con Biotina-U	102

Figura 8	Selección de RNAs por tamaño.	103
Figura 9	Estrategia de doble tinción de los microarrays de embaldosado.	104
Figura 10	Análisis de la dependencia entre densidad de RNAPs activas o nucleosomas con la composición de bases del genoma.	106
Figura 11	Distribución de RNAs nacientes sobre el mapa genómico de levadura.	107
Figura 12	Correlaciones entre los datos de TR calculados por BioGRO y por otras técnicas genómicas.	109
Figura 13	Relación entre la nTR con la longitud de los genes.	112
Figura 14	Efecto de la corrección de los datos de TR para tener en cuenta el efecto del <i>Drop Off</i> .	113
Figura 15	Disminución de las diferencias en las relaciones GRO/DTA y BioGRO/DTA frente a longitud de la ORF tras la corrección de los datos.	114
Figura 16	Perfiles promedio de densidad de RNAP activas.	117
Figura 17	Comparación de perfiles promedio de transcripción naciente con datos de RNA maduro y posicionamiento de nucleosomas.	120
Figura 18	Comparación de perfiles promedio de transcripción naciente con datos de RNAPs totales y posicionamiento de nucleosomas.	122
Figura 19	Perfiles individuales de densidad de RNAPs activas alrededor del TSS.	124
Figura 20	Perfiles promedio de densidad de RNAP activas para distintos grupos de genes.	126
Tabla 3	Niveles de expresión según el tipo de promotor.	127
Figura 21	Comparación entre datos de transcripción naciente y madura alrededor del TSS y del TTS a nivel de genes individuales, ordenados por tamaño.	129
Figura 22	Análisis de la presencia diferencial de RNAPs elongantes más allá del TTS en relación a su tamaño.	131

Figura 23	Aplicación del método BioGRO al estudio de la transcripción naciente en el mutante <i>rpb4Δ</i> .	133
Figura 24	Efecto de la ausencia de Xrn1p en la densidad de RNAPs elongantes a escala genómica.	135
Figura 25	Reanálisis de datos de posicionamiento global de nucleosomas.	138
Figura 26	Relación entre la estructura de la cromatina y la actividad de la RNAPII.	140
Figura 27	Caracterización de la cepa <i>isw2Δ</i> mediante BioGRO.	142
Figura 28	Transcripción naciente promedio de genes influenciados por Isw2 según estudios previos a esta tesis.	143
Figura 29	Estudio comparativo del posicionamiento de nucleosomas alrededor del TSS entre la cepa WT y el mutante <i>isw2Δ</i> .	144
Figura 30	Análisis del cambio de posición de nucleosomas y RNAPs activas en el mutante <i>isw2Δ</i> .	145
Figura 31	Solapamiento entre distintos conjuntos de datos analizados.	146
Tabla 4	Resumen de los mecanismos involucrados en la reactivación de RNAPs paradas.	148
Figura 32	Caracterización de la cepa <i>dst1Δ</i> mediante BioGRO.	149
Figura 33	Defectos en la elongación del mutante <i>dst1Δ</i> en grupos concretos de genes.	150
Figura 34	Estudio del cambio de señal de BioGRO tras un ensayo de inhibición selectiva de RNAPII con α -amanitina.	153
Figura 35	Análisis del metagen 5' tras un ensayo de inhibición selectiva de RNAPII con α -amanitina.	154
Figura 36	Estudio de la transcripción de los genes tRNA con intrones.	156
Figura 37	Estudio de los factores que caracterizan el nivel de expresión de los tRNAs.	158
Figura 38	Comparación de la expresión de las familias de tRNAs en distintas cepas.	160
Figura 39	Efecto de la ausencia de Dst1 o Isw2 en la nTR de los genes tRNA.	162

Figura 40	Estudio de la relación entre la expresión de los tRNAs respecto a la ocurrencia de elementos transponibles en sus proximidades, y sus diferentes orientaciones.	164
Figura 41	Análisis de la relación entre el sesgo en el uso de codones de los tRNAs respecto a la existencia de elementos transponibles en sus proximidades.	164

BIBLIOGRAFÍA

- Acker, J., Conesa, C., y Lefebvre, O. (2013). Yeast RNA polymerase III transcription factors and effectors. *Biochim Biophys Acta*, 1829(3-4), 283–295.
- Al Husini, N., Kudla, P., y Ansari, A. (2013). A Role for CF1A 3' End Processing Complex in Promoter-Associated Transcription. *PLoS genetics*, 9(8), e1003722.
- Alberola, T. M., García-Martínez, J., Antúnez, O., Barceló, A., Ariño, J., y Pérez-ortín, J. E. (2004). A new set of DNA macrochips for the yeast *Saccharomyces cerevisiae*: features and uses. *Int Microbiol*, 7, 199–206.
- Alexander, R., Barrass, J., y Dichtl, B. (2010). RiboSys , a high-resolution , quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3-end processing in *Saccharomyces cerevisiae*. *RNA*, (16), 2570–2580.
- Amaratunga, D., y Cabrera, J. (2001). Analysis of data from viral DNA microchips. *Journal of the American Statistical Association*, 96(456), 1161–1170.
- Anamika, K., Gyenis, À., y Tora, L. (2013). How to stop: the mysterious links among RNA polymerase II occupancy 3' of genes, mRNA 3' processing and termination. *Transcription*, 4(1), 7–12.
- Anderson, J. S., y Parker, R. P. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J*, 17(5), 1497–1506.
- Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., y Lis, J. T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature*, 100, 837–841.
- Ansari, A., y Hampsey, M. (2005). A role for the CPF 3 -end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, 19, 2969–2978.
- Aparicio, O., Geisberg, J. V, Sekinger, E., Yang, A., Moqtaderi, Z., y Struhl, K. (2005). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr Prot Mol Biol*, Chapter 21(3), Unit 21.3.
- Arimbasseri, A. G., y Bhargava, P. (2008). Chromatin structure and expression of a gene transcribed by RNA polymerase III are independent of H2A.Z deposition. *Mol Cel Biol*, 28(8), 2598–2607.
- Armache, K., y Mitterweger, S. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Sci*, 280(8), 7131–7134.
- Baller, J. a, Gao, J., Stamenova, R., Curcio, M. J., y Voytas, D. F. (2012). A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res*, 22(4), 704–713.
- Bartkowiak, B., Liu, P., Phatnani, H., Fuda, N. J., Cooper, J. J., Price, D. H., Greenleaf, A. L. (2010). CDK12 is a transcription elongation- associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev*, 24, 2303–2316.

- Basehoar, A. D., Zanton, S. J., y Pugh, B. F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, 116(5), 699–709.
- Bataille, A. R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Robert, F. (2012). A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Mol Cell*, 45(2), 158–170.
- Bhaumik, S. (2011). Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochim Biophys Acta-Gene Regulatory Mechanisms*, 1809(2), 97–108.
- Bienroth, S., Keller, W., y Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. *EMBO J*, 12(2), 585–594.
- Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. *Cell*, 126(1), 37–47.
- Bolton, E. C., y Boeke, J. D. (2003). Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. *Genome Res.*, 13(2), 254–263.
- Bousquet-Antonelli, C., Presutti, C., y Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell*, 102(6), 765–775.
- Braberg, H., Jin, H., Moehle, E. A., Chan, Y. A., Wang, S., Shales, M., Krogan, N. J. (2013). From Structure to Systems: High-Resolution, Quantitative Genetic Analysis of RNA Polymerase II. *Cell*, 154(4), 775–788.
- Brickner, J. H. (2009). Transcriptional memory at the nuclear periphery. *Current Opinion in Cell Biology*, 21(1), 127–133.
- Brivanlou, A. H., y Darnell, J. E. (2002). Signal transduction and the control of gene expression. *Science*, 295(5556), 813–818.
- Brogaard, K., Xi, L., Wang, J.-P., y Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404), 2-7.
- Bucheli, M. E., y Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. *EMBO J*, 24(12), 2150–2160.
- Buratowski, S. (2003). The CTD code. *Nat Struc and Mol Biol*, 10(9), 679–680.
- Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. *Mol Cell*, 36(4), 541–546.
- Callahan, K. P., y Butler, J. S. (2010). TRAMP Complex Enhances RNA Degradation by the Nuclear Exosome Component Rrp6. *J Biol Chem*, 285(6), 3540–3547.

- Cang, Y., y Prelich, G. (2002). Direct stimulation of transcription by negative cofactor 2 (NC2) through TATA-binding protein (TBP). *PNAS*, 99(20), 12727–12732.
- Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Workman, J. L. (2005). Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*, 123(4), 581–592.
- Chédin, S., Riva, M., y Schultz, P. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev*, 12, 3857–3871.
- Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Wong, E. D. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res*, 40(Database issue), D700–D705.
- Cheung, A. C. M., y Cramer, P. (2012). A movie of RNA polymerase II transcription. *Cell*, 149(7), 1431–1437.
- Chlebowski, A., Lubas, M., Jensen, T. H., y Dziembowski, A. (2013). RNA decay machines: the exosome. *Biochim Biophys Acta*, 1829(6-7), 552–560.
- Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J., y Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev*, 15(24), 3319–3329.
- Chung, H., Dunkel, I., Heise, F., y Linke, C. (2010). The effect of micrococcal nuclease digestion on nucleosome positioning data. *PloS one*, 5(12), e15754.
- Churchman, L. S., y Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330), 368–373.
- Cleary, M. D., Meiering, C. D., Jan, E., Guymon, R., y Boothroyd, J. C. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnology*, 23(2), 232–237.
- Conesa, C., y Acker, J. (2010). Sub1/PC4 a chromatin associated protein with multiple functions in transcription. *RNA biology*, 7(3), 287–290.
- Coppola, J. a, Field, a S., y Luse, D. S. (1983). Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *PNAS*, 80(5), 1251–1255.
- Core, L. J., Waterfall, J. J., y Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909), 1845–1848.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Current opinion in genetics y development*, 14(2), 218–226.

- Creamer, T., Darby, M., y Jamonnak, N. (2011). Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly (A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS genetics*, 7(10), e1002329.
- Curcio, M., y Garfinkel, D. (1992). Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. *Mol Cell Biol*, 12(6), 2813–2825.
- Dahan, N., y Choder, M. (2013). The eukaryotic transcriptional machinery regulates mRNA translation and decay in the cytoplasm. *Biochim Biophys Acta*, 1829(1), 169–173.
- Daniel, J., y Grant, P. (2007). Multi-Tasking on Chromatin with the Saga Coactivator Complexes. *Mutation Research/Fundamental and Molecular*, 618(434), 135–148.
- Daoud, R., Forget, L., y Lang, B. (2012). Yeast mitochondrial RNase P, RNase Z and the RNA degradosome are part of a stable supercomplex. *Nucleic Acids Res*, 40(4), 1728–1736.
- Daulny, A., y Tansey, W. P. (2009). Damage control: DNA repair, transcription, and the ubiquitin-proteasome system. *DNA repair*, 8(4), 444–448.
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., y Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. *PNAS*, 103(14), 5320–5325.
- Jackson, D. A., Iborra, F. J., Manders, E. M. M., y Cook, P. R. (1998). Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei. *Molecular biology of the cell*, 9, 1523–1536.
- Dennis, P. P., Ehrenberg, M., Fange, D., y Bremer, H. (2009). Varying rate of RNA chain elongation during *rrn* transcription in *Escherichia coli*. *Journal of bacteriology*, 191(11), 3740–3746.
- Desai, N., Lee, J., Upadhyay, R., Chu, Y., Moir, R. D., y Willis, I. M. (2005). Two steps in Maf1-dependent repression of transcription by RNA polymerase III. *J Biol Chem*, 280(8), 6455–6462.
- Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., y Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends in genetics : TIG*, 23(12), 614–622.
- Dikstein, R. (2011). The unexpected traits associated with core promoter elements. *Transcription*, 2(5), 201–206.
- Dittmar, K., Goodenbour, J. M., y Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS genetics*, 2(12), e221.
- Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C. C., Zimmer, R., Mages, J., Koszinowski, U. H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9), 1959–1972.

- Donze, D. (2012). Extra-transcriptional functions of RNA Polymerase III complexes: TFIIC as a potential global chromatin bookmark. *Gene*, 493(2), 169–175.
- Dos Reis, M., Savva, R., y Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res*, 32(17), 5036–5044.
- Drouin, S., Laramée, L., Jacques, P.-É., Forest, A., Bergeron, M., y Robert, F. (2010). DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS genetics*, 6(10), e1001173.
- Drummond, D. A., y Wilke, C. O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, 134(2), 341–352.
- Durand, S., y Lykke-Andersen, J. (2011). SnapShot: Nonsense-mediated mRNA decay. *Cell*, 145(2), 324e1–324.e2.
- Edwards, A. M., Kane, C. M., Young, R. A., y Kornberg, R. D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem*, 266(1), 71–75.
- Egloff, S., O'Reilly, D., y Chapman, R. (2007). Serine 7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, 318(5857), 1777–1779.
- Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., y Murphy, S. (2012). Ser7 phosphorylation of the CTD recruits the RPA2 Ser5 phosphatase to snRNA genes. *Mol Cell*, 45(1), 111–122.
- Ehrensberger, A. H., Kelly, G. P., & Svejstrup, J. Q. (2013). Mechanistic Interpretation of Promoter-Proximal Peaks and RNAPII Density Maps. *Cell*, 154(4), 713–715.
- Ezeokonkwo, C., Ghazy, M. a, Zhelkovsky, A., Yeh, P.-C., y Moore, C. (2012). Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in *Saccharomyces cerevisiae*. *FEBS letters*, 586(8), 1173–1178.
- Fernández-Tornero, C., Böttcher, B., Rashid, U. J., Steuerwald, U., Flörchinger, B., Devos, D. P., Müller, C. (2010). Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J*, 29, 3762–3772.
- Fish, R., y Kane, C. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Bioch Biophys Acta*, 1577(2002), 287–307.
- Formosa, T., Eriksson, P., Wittmeyer, J., Ginn, J., Yu, Y., y Stillman, D. J. (2001). Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *EMBO J*, 20(13), 3506–3517.
- Friedel, C. C., y Dölken, L. (2009). Metabolic tagging and purification of nascent RNA: implications for transcriptomics. *Mol BioSys*, 5(11), 1271–1278.

- Fromm, G., Gilchrist, D. a., y Adelman, K. (2013). SnapShot: Transcription Regulation: Pausing. *Cell*, 153(4), 930–930.e1.
- Fuda, N. J., y Lis, J. T. (2013). A new player in Pol II pausing. *EMBO J*, 32(13), 1796–1798.
- Fujimura, T., y Esteban, R. (2010). Yeast double-stranded RNA virus L-A deliberately synthesizes RNA transcripts with 5'-diphosphate. *J Biol Chem*, 285(30), 22911–22918.
- García, A., Rosonina, E., Manley, J. L., y Calvo, O. (2010). Sub1 globally regulates RNA polymerase II C-terminal domain phosphorylation. *Mol Cell Biol*, 30(21), 5180–5193.
- García-Martínez, J., Aranda, A., y Pérez-Ortín, J. E. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell*, 15(2), 303–313.
- García-Martínez, J., Pelechano, V., y Pérez-Ortín, J. E. (2011). Genomic-Wide Methods to Evaluate Transcription Rates in Yeast. (A. Becskei, Ed.) *Yeast Genetic Networks: Methods and Protocols, Methods in Molecular Biology*, 734, 25–44.
- Garneau, N. L., Wilusz, J., y Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat Reviews*, 8(2), 113–126.
- Gaston, K., y Jayaraman, P.-S. (2003). Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cellular and Molecular Life Sciences (CMLS)*, 60(4), 721–741.
- Gelbart, M., Bachman, N., Delrow, J., Boeke, J. D., y Tsukiyama, T. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev*, 19, 942–954.
- Ghavi-Helm, Y., Michaut, M., Acker, J., Aude, J.-C., Thuriaux, P., Werner, M., y Soutourina, J. (2008). Genome-wide location analysis reveals a role of TFIIIS in RNA polymerase III transcription. *Genes Dev*, 22(14), 1934–1947.
- Gnatt, a L., Cramer, P., Fu, J., Bushnell, D. a, y Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 292(5523), 1876–1882.
- Goffeau, A., Barrell, B., Bussey, H., y Davis, R. (1996). Life with 6000 genes. *Science*, 274, 546–567.
- Goldstrohm, A. C., y Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nat Reviews*, 9(4), 337–344.
- Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., y Choder, M. (2008). Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev*, 22(15), 2022–2027.

- Gómez-Herreros, F., de Miguel-Jiménez, L., Millán-Zambrano, G., Peñate, X., Delgado-Ramos, L., Muñoz-Centeno, M. C., & Chávez, S. (2012). One step back before moving forward: Regulation of transcription elongation by arrest and backtracking. *FEBS Letters*, *586*(18), 2820–2825.
- Görnemann, J., Kotovic, K. M., Hujer, K., y Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell*, *19*(1), 53–63.
- Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K., y Hinnebusch, A. G. (2007). Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol Cell*, *25*(1), 31–42.
- Graczyk, D., Debski, J., Muszyńska, G., Bretner, M., Lefebvre, O., y Boguta, M. (2011). Casein kinase II-mediated phosphorylation of general repressor Maf1 triggers RNA polymerase III activation. *PNAS*, *108*(12), 4926–4931.
- Grantham, R., Gautier, C., y Gouy, M. (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*, *8*(1), 49–62.
- Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D. y Hughes, T. R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol*, *24*(12), 5534–5547.
- Grosso, A., Almeida, S. de, Braga, J., y Carmo-Fonseca, M. (2012). Dynamic transitions in RNA polymerase II density profiles during transcription termination. *Genome Res*, *22*, 1447–1456.
- Gu, W., Jackman, J. E., Lohan, A. J., Gray, M. W., y Phizicky, E. M. (2003). tRNA^{His} maturation: an essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNA^{His}. *Genes Dev*, *17*(23), 2889–2901.
- Guo-Cheng, Y., Yuen-Jong, L., Dion, M. F., Slack, M. D., Wu, L. F., y Altschuler, S. J. (2005). Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae*. *PNAS*, *102*(26), 9294–9299.
- Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., y Kraus, W. L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, *145*(4), 622–634.
- Haimovich, G., Medina, D. A., Causse, S. Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J. E., Darzack, X., y Choder, M. (2013). Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis. *Cell*, *153*(5), 1000–1011.
- Hampsey, M., Singh, B. N., Ansari, A., Lainé, J.-P., y Krishnamurthy, S. (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. *Advances in enzyme regulation*, *51*(1), 118–125.

- Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., y Choder, M. (2010). RNA Polymerase II Subunits Link Transcription and mRNA Decay to Translation. *Cell*, 143(4), 552–563.
- Hartley, P., y Madhani, H. (2009). Mechanisms that specify nucleosome location and identity. *Cell*, 137, 445–458.
- Hausmann, S., y Shuman, S. (2002). Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem*, 277(24), 21213–21220.
- Hayles, B., Yellaboina, S., y Wang, D. (2010). Comparing transcription rate and mRNA abundance as parameters for biochemical pathway and network analysis. *PLoS one*, 5(3), e9908.
- Helden, J. Van, Olmo, M. I. del, y Pérez-Ortín, J. E. (2000). Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic Acids Res*, 28(4), 1000–1010.
- Hennig, B. P., Bendrin, K., Zhou, Y., y Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO reports*, 13(11), 997–1003.
- Hirayoshi, K., y Lis, J. T. (1999). Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases. *Chromatin*, 304, 351–362.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., y Bustamante, C. (2009). Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II. *Science*, 325(5940), 626–628.
- Hopper, A. K. (2013). Transfer RNA Post-Transcriptional Processing, Turnover, and Subcellular Dynamics in the Yeast *Saccharomyces cerevisiae*. *Genetics*, 194(1), 43–67.
- Hota, S. K., y Bartholomew, B. (2011). Diversity of operation in ATP-dependent chromatin remodelers. *Biochim Biophys Acta*, 1809(9), 476–487.
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., y Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), 218–223.
- Jackson, D. a, Hassan, a B., Errington, R. J., y Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J*, 12(3), 1059–1065.
- Jasiak, A. J., Armache, K.-J., Martens, B., Jansen, R.-P., y Cramer, P. (2006). Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell*, 23(1), 71–81.
- Jiang, C., y Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, 10(3), 161–172.

- Jimeno-González, S., Gómez-Herreros, F., Alepuz, P. M., y Chávez, S. (2006). A gene-specific requirement for FACT during transcription is related to the chromatin organization of the transcribed region. *Mol Cell Biol*, 26(23), 8710–8721.
- Jin, J., Bai, L., Johnson, D. S., Fulbright, R. M., Kireeva, M. L., Kashlev, M., y Wang, M. D. (2010). Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nat Structural and Molecular Biology*, 17(6), 745–752.
- Jones, B. (2012). Chromatin: A model for nucleosome positioning. *Nat Rev Genet*, 13(10), 674–675.
- Kadonaga, J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, 116(2), 247–257.
- Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., Reid, G. (2008). Transient cyclical methylation of promoter DNA. *Nature*, 452(7183), 112–115.
- Kaplan, C. D., Laprade, L., y Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, 301(5636), 1096–1099.
- Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, 458(7236), 362–366.
- Kassabov, S. R., Zhang, B., Persinger, J., y Bartholomew, B. (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell*, 11(2), 391–403.
- Kassavetis, G. a, y Steiner, D. F. (2006). Nhp6 is a transcriptional initiation fidelity factor for RNA polymerase III transcription in vitro and in vivo. *J Biol Chem*, 281(11), 7445–7451.
- Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., y Proudfoot, N. J. (2008). Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev*, 22(8), 1082–1092.
- Kazerouninia, A., Ngo, B., y Martinson, H. G. (2010). Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. *RNA*, 16(1), 197–210.
- Kenzelmann, M., Maertens, S., Hergenahn, M., Kueffer, S., Hotz-Wagenblatt, A., Li, L., Schütz, G. (2007). Microarray analysis of newly synthesized RNA in cells and animals. *PNAS*, 104(15), 1–6.
- Keogh, M., Podolny, V., y Buratowski, S. (2003). Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol*, 23(19), 7005–7018.

- Khodor, Y., Rodriguez, J., Abruzzi, K. C., Tang, C. H. A., Marr, M. T., y Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev*, *25*, 2502–2512.
- Kim, J., Vanguri, S., y Boeke, J. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome. *Genome Res*, *8*, 464–478.
- Kim, J.-H., Saraf, A., Florens, L., Washburn, M., y Workman, J. L. (2010). Gcn5 regulates the dissociation of SWI/SNF from chromatin by acetylation of Swi2/Snf2. *Genes Dev*, *24*(24), 2766–2771.
- Kim, K.-Y., y Levin, D. E. (2011). Mpk1 MAPK Association with the Paf1 Complex Blocks Sen1-Mediated Premature Transcription Termination. *Cell*, *144*(5), 745–756.
- Kim, T., y Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell*, *137*(2), 259–272.
- Kim, T. S., Liu, C. L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., y Rando, O. J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. *Genome Biology*, *11*(7), R75.
- Kinsey, P. T., y Sandmeyer, S. B. (1991). Adjacent pol II and pol III promoters: transcription of the yeast retrotransposon Ty3 and a target tRNA gene. *Nucleic Acids Res*, *19*(6), 1317–1324.
- Kireeva, M. L., Hancock, B., Cremona, G. H., Walter, W., Studitsky, V. M., y Kashlev, M. (2005). Nature of the Nucleosomal Barrier to RNA Polymerase II. *Mol Cell*, *18*(1), 97–108.
- Koerber, R. T., Rhee, H. S., Jiang, C., y Pugh, B. F. (2009). Interaction of Transcriptional Regulators with Specific Nucleosomes across the *Saccharomyces* Genome. *Mol Cell*, *35*(6), 889–902.
- Komarnitsky, P., Cho, E., y Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, *14*, 2452–2460.
- Kostrewa, D., Zeller, M. E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., y Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, *462*(7271), 323–330.
- Kraeva, R. I., Krastev, D. B., Roguev, A., Ivanova, A., Nedelcheva-Velva, M. N., y Stoynov, S. S. (2007). Stability of mRNA/DNA and DNA/DNA duplexes affects mRNA transcription. *PLoS one*, *2*(3), e290.
- Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., y Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol cell*, *14*(3), 387–94.

- Krogan, N., y Greenblatt, J. (2001). Characterization of a Six-Subunit Holo-Elongator Complex Required for the Regulated Expression of a Group of Genes in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 21(23), 8203–8212.
- Kruesi, W. S., Core, L. J., Waters, C. T., Lis, J. T., y Meyer, B. J. (2013). Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *eLife*, 2(Pol II), e00808–e00808.
- Kruk, J. A., Dutta, A., Fu, J., Gilmour, D. S., y Reese, J. C. (2011). The multifunctional Ccr4 – Not complex directly promotes transcription elongation. *Genes Dev*, 25, 581–593.
- Kuehner, J. N., Pearson, E. L., y Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Reviews*, 12(5), 1–12.
- Kulaeva, O. I., Gaykalova, D. a, Pestov, N. a, Golovastov, V. V, Vassylyev, D. G., Artsimovitch, I., y Studitsky, V. M. (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat Structural and Molecular Biology*, 16(12), 1272–1278.
- Kumar, Y., y Bhargava, P. (2013). A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. *BMC Genomics*, 14(1), 402.
- Kwak, H., Fuda, N. J., Core, L. J., y Lis, J. T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 339(6122), 950–3.
- Lacadie, S. a, Tardiff, D. F., Kadener, S., y Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev*, 20(15), 2055–2066.
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., y Tollervey, D. (2005). RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. *Cell*, 121(5), 713–724.
- Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., y Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, 39(10), 1235–1244.
- Lin, P., Marshall, N., y Dahmus, M. (2002). CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. *Progress in nucleic acid research and Molecular Biology*, 72, 333–365.
- Lorenzen, K., Vannini, A., Cramer, P., y Heck, A. J. R. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure*, 15(10), 1237–1245.
- Lunde, B., Reichow, S., y Kim, M. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Structural and Molecular Biology*, 17(10), 1195–1201.

- Luse, D. S., y Studitsky, V. M. (2011). The mechanism of nucleosome traversal by RNA polymerase II: roles for template uncoiling and transcript elongation factors. *RNA Biol*, 8(4), 581–585.
- Man, O., y Pilpel, Y. (2007). Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat Genet*, 39(3), 415–421.
- Mandal, S., Cho, H., y Kim, S. (2002). FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. *Mol Cell Biol*, 22(21), 7543–7552.
- Mannervik, M., Nibu, Y., Zhang, H., y Levine, M. (1999). Transcriptional Coregulators in Development. *Science*, 284(5414), 606–609.
- Mao, C., Brown, C. R., Griesenbeck, J., y Boeger, H. (2011). Occlusion of regulatory sequences by promoter nucleosomes in vivo. *PloS one*, 6(3), e17521.
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M., y Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J*, 30(7), 1302–1310.
- Mason, P. B., y Struhl, K. (2005). Distinction and Relationship between Elongation Rate and Processivity of RNA Polymerase II In Vivo. *Mol Cell*, 17(6), 831–840.
- Mavrich, T. N., Ioshikhes, I. P., Venters, B. J., Jiang, C., Tomsho, L. P., Qi, J., Pugh, B. F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*, 18(7), 1073–1083.
- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., y Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Structural and Molecular Biology*, 17(10), 1272–1278.
- McKinlay, A., Araya, C. L., y Fields, S. (2011). Genome-Wide Analysis of Nascent Transcription in *Saccharomyces cerevisiae*. G3: *Genes, Genomes, Genetics*, 1(7), 549–558.
- Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., y Cramer, P. (2005). A structural perspective of CTD function. *Genes Dev*, 19(12), 1401–1415.
- Mellor, J., y Morillon, A. (2004). ISWI complexes in *Saccharomyces cerevisiae*. *Biochim Biophys Acta*, 1677(1-3), 100–112.
- Melvin, W.T., Milne, H.B., Slater, A.A., y Allen, H.J. (1978). Incorporation of 6-Thioguanosine and 4-Thiouridine into RNA Application to Isolation of Newly Synthesised RNA by Affinity Chromatography. *European Journal of Biochemistry*, 379, 373–379.
- Miguel, A., Montón, F., Li, L., Gómez-Herreros, F., Chávez, S., Alepuz, P., y Pérez-Ortín, J. E. (2013). External conditions inversely change the RNA polymerase II

- elongation rate and density in yeast. *Biochim Biophys Acta, Gene Regul Mechan* (en prensa).
- Miller, C., Schwalb, B., Ouml rn, Maier, K., Schulz, D., mcke, S. D. uuml, Zacher, B., Cramer, P. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Sys Biol*, 7, 1–13.
- Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A., y Tollervey, D. (2008). A Yeast Exosome Cofactor, Mpp6, Functions in RNA Surveillance and in the Degradation of Noncoding RNA Transcripts. *Mol Cel Biol*, 28(17), 5446–5457.
- Mischo, H. E., y Proudfoot, N. J. (2012). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta*, 1829(1), 174–185.
- Mohrmann, L., y Verrijzer, C. P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta*, 1681(2-3), 59–73.
- Montel, F., y Castelnovo, M. (2011). RSC remodeling of oligo-nucleosomes: an atomic force microscopy study. *Nucleic Acids Res*, 39(7), 2571–2579.
- Moqtaderi, Z., Geisberg, J. V., Jin, Y., Fan, X., y Struhl, K. (2013). Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. *PNAS*, 110(27), 11073–11078.
- Morillon, A., y Bénard, L. (2002). Differential Effects of Chromatin and Gcn4 on the 50-Fold Range of Expression among Individual Yeast Ty1 Retrotransposons. *Mol Cel Biol*, 22(7), 2078–2088.
- Mosley, A. L., Pattenden, S. G., Carey, M., Venkatesh, S., Gilmore, J. M., Florens, L., Washburn, M. P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol cell*, 34(2), 168–178.
- Muhlrad, D., Decker, C. J., y Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes Dev*, 8(7), 855–866.
- Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S. J., y Boeke, J. D. (2012). Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. *Genome Res*, 22(4), 693–703.
- Müller, H.-M., Kenny, E. E., y Sternberg, P. W. (2004). Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS biology*, 2(11), e309.

- Ng, H. H., Robert, F., Young, R. a, y Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*, *11*(3), 709–719.
- O'Sullivan, J. M., Tan-Wong, S. M., Morillon, A., Lee, B., Coles, J., Mellor, J., y Proudfoot, N. J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nat Genetics*, *36*(9), 1014–1018.
- Oesterreich, F. C., Bieberstein, N., y Neugebauer, K. M. (2011). Pause locally, splice globally. *Trends in cell biology*, *21*(6), 328–335.
- Oesterreich, F. C., Preibisch, S., y Neugebauer, K. M. (2010). Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. *Mol Cell*, *40*(4), 571–581.
- Ohtsu, M., Kawate, M., Fukuoka, M., Gunji, W., Hanaoka, F., Utsugi, T., Murakami, Y. (2008). Novel DNA Microarray System for Analysis of Nascent mRNAs. *DNA Res*, *15*(4), 241–251.
- Orphanides, G., y Reinberg, D. (2002). A unified theory of gene expression. *Cell*, *108*, 439–451.
- Ozanick, S. G., Wang, X., Costanzo, M., Brost, R. L., Boone, C., y Anderson, J. T. (2009). Rex1p deficiency leads to accumulation of precursor initiator tRNAMet and polyadenylation of substrate RNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, *37*(1), 298–308.
- Parker, R. (2012). RNA degradation in *Saccharomyces cerevisiae*. *Genetics*, *191*(3), 671–702.
- Paule, M. R., y White, R. J. (1988). Transcription by RNA polymerase III. *Annual review of Biochemistry*, *28*(6), 1283–1298.
- Pei, Y., Schwer, B., y Shuman, S. (2003). Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J Biol Chem*, *278*(9), 7180–7188.
- Peiro-Chova, L., y Estruch, F. (2007). Specific Defects in Different Transcription Complexes Compensate for the Requirement of the Negative Cofactor 2 Repressor in *Saccharomyces cerevisiae*. *Genetics*, *176*(1), 125–138.
- Pelechano, V, Chávez, S., y Pérez-Ortín, J. (2010). A complete set of nascent transcription rates for yeast genes. *PLoS One*, *5*(11), e15442.
- Pelechano, Vicent, Jimeno-González, S., Rodríguez-Gil, A., García-Martínez, J., Pérez-Ortín, J. E., y Chávez, S. (2009). Regulon-Specific Control of Transcription Elongation across the Yeast Genome. *PLoS Genetics*, *5*(8), e1000614.
- Pelechano, Vicent, y Pérez-Ortín, J. E. (2010). There is a steady-state transcriptome in exponentially growing yeast cells. *Yeast*, *27*(7), 413–422.

- Pelechano, Vicent, Wei, W., y Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, 497(5), 127–131.
- Pelham, H. R. B. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *PNAS*, 77(7), 4170–4174.
- Perales, R., y Bentley, D. (2009). Cotranscriptionality: The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol Cell*, 36(2), 178–191.
- Percudani, R., Pavesi, A., y Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol*, 268, 322–330.
- Pérez-Ortín, J. E., Alepuz, P., Chávez, S., y Choder, M. (2013). Eukaryotic mRNA Decay: Methodologies, Pathways, and Links to Other Stages of Gene Expression. *J Mol Biol*, (In press), 1–26.
- Pérez-Ortín, J. E., Jordán-Pla, A., y Pelechano, V. (2011). A genomic view of mRNA turnover in yeast. *Comptes Rendus Biologies*, 334, 647–654.
- Perez-Ortin, J. E., Matallana, E., y Franco, L. (1989). Chromatin structure of yeast genes. *Yeast*, 5(4), 219–238.
- Peterson, C. L., y Workman, J. L. (2000). Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Current opinion in genetics and development*, 10(2), 187–192.
- Phatnani, H. P., y Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev*, 20(21), 2922–2936.
- Phizicky, E., y Hopper, A. (2010). tRNA biology charges to the front. *Genes Dev*, 24, 1832–1860.
- Pokholok, D., Hannett, N., y Young, R. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol cell*, 9(4), 799–809.
- Prather, D., Larschan, E., y Winston, F. (2005). Evidence that the Elongation Factor TFIIS Plays a Role in Transcription Initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol Cel Biol*, 25(7), 2650–2659.
- Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast*, 16(8), 675–729.
- Proudfoot, N. (2011). Ending the message: poly (A) signals then and now. *Genes Dev*, 25, 1770–1782.
- Qi, X., Daily, K., Nguyen, K., Wang, H., Mayhew, D., Rigor, P., Sandmeyer, S. (2012). Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Res*, 22(4), 681–692.

- Rando, O. J., y Winston, F. (2012). Chromatin and transcription in yeast. *Genetics*, 190(2), 351–387.
- Raveh-Sadka, T., Levo, M., y Segal, E. (2009). Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res*, 19(8), 1480–1496.
- Rhee, H. S., y Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483(7389), 1–7.
- Richard, P., y Manley, J. L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev*, 23(11), 1247–1269.
- Richmond, T. J., y Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature*, 423(6936), 145–150.
- Rodriguez-Gil, A., Garcia-Martinez, J., Pelechano, V., Munoz-Centeno, M. d l C., Geli, V., Perez-Ortin, J. E., Chávez, S. (2010). The distribution of active RNA polymerase II along the transcribed region is gene-specific and controlled by elongation factors. *Nucleic Acids Res*, 38(14), 4651–4664.
- Rondón, A. G., Mischo, H. E., Kawauchi, J., y Proudfoot, N. J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in *S. cerevisiae*. *Mol Cell*, 36(1), 88–98.
- Rudd, M., y Luse, D. (1996). Amanitin Greatly Reduces the Rate of Transcription by RNA Polymerase II Ternary Complexes but Fails to Inhibit Some Transcript Cleavage Modes. *J Biol Chem*, 271(15), 21549–21558.
- Ryan, K., Calvo, O., y Manley, J. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, 10, 565–573.
- Saeki, H., y Svejstrup, J. Q. (2009). Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. *Mol Cell*, 35(2), 191–205.
- Saunders, A., Core, L. J., y Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nat Reviews. Molecular cell biology*, 7(8), 557–567.
- Schmid, M., Poulsen, M. B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L. M., Moore, C., y Jensen, T. H. (2012). Rrp6p Controls mRNA Poly(A) Tail Length and Its Decoration with Poly(A) Binding Proteins. *Mol Cell*, 47(7), 1–14.
- Schoenberg, D., y Maquat, L. (2012). Regulation of cytoplasmic mRNA decay. *Nat Reviews Genetics*, 13(4), 246–259.
- Schramm, L., y Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev*, 16(20), 2593–2620.
- Segal, E., y Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat Reviews Genetics*, 10(7), 443–456.

- Seizl, M., Hartmann, H., Hoeg, F., Kurth, F., Martin, D. E., Söding, J., y Cramer, P. (2011). A Conserved GA Element in TATA-Less RNA Polymerase II Promoters. *PLoS ONE*, 6(11), e27595.
- Selth, L. A., Sigurdsson, S., y Svejstrup, J. Q. (2010). Transcript Elongation by RNA Polymerase II. *Annual Review of Biochemistry*, 79(1), 271–293.
- Shalem, O., Groisman, B., Choder, M., Dahan, O., y Pilpel, Y. (2011). Transcriptome Kinetics Is Governed by a Genome-Wide Coupling of mRNA Production and Degradation: A Role for RNA Pol II. *PLoS genetics*, 7(9), e1002273.
- Shandilya, J., y Roberts, S. G. E. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta*, 1819(5), 391–400.
- Sharp, P., y Li, W. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15(3), 1281–1295.
- Sigurdsson, S., Dirac-Svejstrup, A. B., y Svejstrup, J. Q. (2010). Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Mol Cell*, 38(2), 202–210.
- Sikorski, T. W., y Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology*, 21(3), 344–351.
- Sims, R. J., Belotserkovskaya, R., y Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev*, 18(20), 2437–2468.
- Smale, S. T., y Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual review of biochemistry*, 72, 449–479.
- Somesh, B. P., Sigurdsson, S., Saeki, H., Erdjument-Bromage, H., Tempst, P., y Svejstrup, J. Q. (2007). Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. *Cell*, 129(1), 57–68.
- Somesh, B., Reid, J., y Liu, W. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell*, 121, 913–923.
- Spåhr, H., Calero, G., Bushnell, D. a, y Kornberg, R. D. (2009). Schizosacharomyces pombe RNA polymerase II at 3.6-Å resolution. *PNAS*, 106(23), 9185–9190.
- Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., y Brow, D. a. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol cell*, 24(5), 735–746.
- Strahl, B., Grant, P., y Briggs, S. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Mol Cell Biol*, 22(5), 1298–1306.

- Studitsky, V. M., y George A. Kassavetis, E. Peter Geiduschek, G. F. (1997). Mechanism of Transcription Through the Nucleosome by Eukaryotic RNA Polymerase. *Science*, 278(5345), 1960–1963.
- Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Cramer, P. (2012). Comparative Dynamic Transcriptome Analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res*, 22(7), 1350–1359.
- Taatjes, D. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in biochemical sciences*, 35(6), 315–322.
- Tan-Wong, S. M., Zaugg, J. B., Camblong, J., Xu, Z., Zhang, D. W., Mischo, H. E., Proudfoot, N. J. (2012). Gene Loops Enhance Transcriptional Directionality. *Science*, 338(6107), 671–675.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, 6(5), 377–382.
- Tardiff, D. F., Lacadie, S. A., y Rosbash, M. (2007). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly post-transcriptional. *Mol Cell*, 24(6), 917–929.
- Tavenet, A., Suleau, A., Dubreuil, G., Ferrari, R., Ducrot, C., Michaut, M., Acker, J. (2009). Genome-wide location analysis reveals a role for Sub1 in RNA polymerase III transcription. *PNAS*, 106(34), 14265–14270.
- Teichmann, M., Dieci, G., Pascali, C., y Boldina, G. (2010). General transcription factors and subunits of RNA polymerase III: Paralogs for promoter- and cell type-specific transcription in multicellular eukaryotes. *Transcription*, 1(3), 130–135.
- Terzi, N., Churchman, L. S., Vasiljeva, L., Weissman, J., y Buratowski, S. (2011). H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol Cell Biol*, 31(17), 3569–3583.
- Thompson, M., Haeusler, R. a, Good, P. D., y Engelke, D. R. (2003). Nucleolar clustering of dispersed tRNA genes. *Science*, 302(5649), 1399–1401.
- Tillo, D., y Hughes, T. R. (2009). G+C content dominates intrinsic nucleosome occupancy. *BMC bioinformatics*, 10, 442.
- Tirosh, I., Sigal, N., y Barkai, N. (2010). Widespread remodeling of mid-coding sequence nucleosomes by Isw1. *Genome Biology*, 11(5), R49.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2), 344–354.

- Udugama, M., Sabri, A., y Bartholomew, B. (2011). The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor. *Mol Cell Biol*, 31(4), 662–673.
- Van Dijk, E. L., Chen, C. L., D'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., D'Aubenton-Carafa, Y. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, 475(7), 114–117.
- Van Vugt, J. J. F. A., Raney, M., Campsteijn, C., y Logie, C. (2007). The ins and outs of ATP-dependent chromatin remodeling in budding yeast: biophysical and proteomic perspectives. *Biochim Biophys Acta*, 1769(3), 153–171.
- Vannini, A. (2013). A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries. *Biochim Biophys Acta*, 1829(3-4), 258–264.
- Vannini, A., y Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45(4), 439–446.
- Vignali, M., y Hassan, A. (2000). ATP-Dependent Chromatin-Remodeling Complexes. *Mol Cell Biol*, 20(6), 1899–1910.
- Voliotis, M., Cohen, N., Molina-París, C., y Liverpool, T. B. (2008). Fluctuations, Pauses, and Backtracking in DNA Transcription. *Biophysj*, 94(2), 334–348.
- Wahl, M. C., Will, C. L., y Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, 136(4), 701–718.
- Waldminghaus, T., y Skarstad, K. (2010). ChIP on Chip: surprising results are often artifacts. *BMC Genomics*, 11(1), 414–429.
- Wang, Y., y Liu, C. (2002). Precision and functional specificity in mRNA decay. *PNAS*, 99(9), 5860–5865.
- Wang, Z., y Roeder, R. G. (1998). DNA topoisomerase I and PC4 can interact with human TFIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III. *Mol cell*, 1(5), 749–757.
- Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences*, 24(11), 437–440.
- Weake, V. M., y Workman, J. L. (2010). Inducible gene expression: diverse regulatory mechanisms. *Nat Reviews Genetics*, 11(6), 426–437.
- Wei, W., Pelechano, V., Järvelin, A. I., y Steinmetz, L. M. (2011). Functional consequences of bidirectional promoters. *TIG*, 27(7), 267–276.
- Weiner, A. M., Deininger, P. L., y Efstratiadis, A. (1986). Nonviral Retroposons: Genes, Pseudogenes, and Transposable Elements Generated by the Reverse Flow of Genetic Information. *Annual Review of Biochemistry*, 55(1), 631–661.

- Whitehouse, I, Flaus, A., Cairns, B. R., White, M. F., Workman, J. L., y Owen-Hughes, T. (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*, 400(6746), 784–787.
- Whitehouse, Iestyn, Rando, O. J., Delrow, J., y Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, 450(7172), 1031–5.
- Wilkening, S., Pelechano, V., Järvelin, A. I., Tekkedil, M. M., Anders, S., Benes, V., y Steinmetz, L. M. (2013). An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res*, 41(5), e65.
- Willis, I. M., Desai, N., y Upadhyay, R. (2004). Signaling Repression of Transcription by RNA Polymerase III in Yeast. *Progress in Nucleic Acid Research and Molecular Biology*, 77, 323–353.
- Wittmeyer, J., y Formosa, T. (1997). The *Saccharomyces cerevisiae* DNA polymerase alpha catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. *Mol Cell Biol*, 17(7), 4178–4190.
- Woychik, N. a, Lane, W. S., y Young, R. a. (1991). Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem*, 266(28), 19053–19055.
- Woychik, N. a, y Young, R. a. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol Cell Biol*, 9(7), 2854–1859.
- Wu, P., y Winston, F. (2002). Analysis of Spt7 Function in the *Saccharomyces cerevisiae* SAGA Coactivator Complex. *Mol Cell Biol*, 22(15), 5367–5379.
- Wullschleger, S., Loewith, R., & Hall, M. N. (2006). TOR signaling in growth and metabolism. *Cell*, 124(3), 471–84.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Séraphin, B. (2005). Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell*, 121(5), 725–737.
- Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232), 1033–1037.
- Yang, Q., y Doublé, S. (2011). Structural biology of poly(A) site definition. *Wiley Interdisciplinary Reviews: RNA*, 2(5), 732–747.
- Yassour, M., Kaplan, T., Jaimovich, A., y Friedman, N. (2008). Nucleosome positioning from tiling microarray data. *Bioinformatics*, 24(13), i139–i146.
- Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T., y Pugh, B. F. (2012). Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. *Cell*, 149(7), 1461–1473.

- Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., y Jones, K. a. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev*, 21(2), 160–174.
- Yoshihisa, T., y Yunoki-Esaki, K. (2003). Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. *Mol Biol Cell*, 14(August), 3266–3279.
- Zacher, B., Torkler, P., y Tresch, A. (2011). Analysis of Affymetrix ChIP-Chip Data Using Starr and R/Bioconductor. *Cold Spring Harbor Protocols*, 2011(5), pdb.top110–pdb.top110.
- Zawel, L., Kumar, K. P., y Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev*, 9(12), 1479–1490.
- Zhang, Y., Moqtaderi, Z., Rattner, B. P., Euskirchen, G., Snyder, M., Kadonaga, J. T., Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Structural and Molecular Biology*, 16(8), 847–852.
- Zhang, Z., Wippo, C. J., Wal, M., Ward, E., Korber, P., y Pugh, B. F. (2011). A Packing Mechanism for Nucleosome Organization Reconstituted Across a Eukaryotic Genome. *Science*, 332(6032), 977–980.
- Zhang, Zhihong, y Dietrich, F. S. (2005). Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res*, 33(9), 2838–2851.