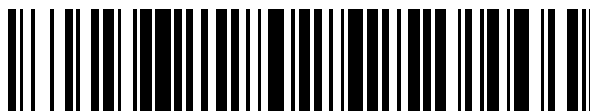


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 785 274**

51 Int. Cl.:

C12N 15/63 (2006.01)

C12N 15/113 (2010.01)

A61K 38/16 (2006.01)

C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.07.2013 PCT/US2013/049451**

87 Fecha y número de publicación internacional: **09.01.2014 WO14008474**

96 Fecha de presentación y número de la solicitud europea: **05.07.2013 E 13813188 (3)**

97 Fecha y número de publicación de la concesión europea: **25.12.2019 EP 2866822**

54 Título: **La RNPSn U1 regula la expresión génica y modula la oncogenicidad**

30 Prioridad:

05.07.2012 US 201261668168 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.10.2020

73 Titular/es:

**THE TRUSTEES OF THE UNIVERSITY OF
PENNSYLVANIA (100.0%)
Center for Technology Transfer, 3160 Chestnut
Street, Suite 200
Philadelphia PA 19104-6283, US**

72 Inventor/es:

**DREYFUSS, GIDEON;
WAN, LILI;
OH, JUNG MIN;
PINTO, ANNA MARIA;
YOUNIS, IHAB y
BERG, MICHAEL, GEORGE**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 785 274 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

La RNPsn U1 regula la expresión génica y modula la oncogenicidad

Antecedentes de la invención

Los ARN mensajeros en las células eucariotas son producidos a partir de transcritos precursores (pre-ARNm) por procesamiento posterior a la transcripción. En los metazoos, dos reacciones de procesamiento son particularmente extensas y contribuyen lo más significativamente a la diversidad del transcriptoma de ARNm - corte y empalme de intrones y escisión y poliadenilación alternativas (Di Giammartino et al., 2011, Mol Cell 43:853-866; Hartmann y Valcarcel, 2009, Curr Opin Cell Biol 21:377-386; Wang et al., 2008, Nature 456:470-476). El corte y empalme se realiza por un espliceosoma que se ensambla en cada intrón y está comprendido predominantemente por RNP nucleares pequeñas (RNPsn), RNPsn U1, U2, U4, U5 y U6, en la misma estequiometría (Nilsen, 2003, Bioessays 25(12):1147-1149; Wahl et al., 2009, Cell 136:701-718). La RNPsn U1 (U1) juega un papel esencial en la definición del sitio de corte y empalme en 5' (ss) por emparejamiento de bases ARN:ARN a través de una secuencia de nueve nucleótidos (nt) en 5' del ARNsn de U1. Usando el oligonucleótido morfolino antisentido complementario al extremo 5' del ARNsn de U1 (AMO de U1) que interfiere con la función de la RNPsn U1 en las células humanas, se observó la acumulación de intrones en muchos transcritos, como se espera por la inhibición del corte y empalme (Kaida et al., 2010, Nature 468, 664-668). Sin embargo, además, la mayoría de los pre-ARNm terminaron prematuramente a partir de PAS crípticos en intrones, típicamente a una distancia corta del sitio de inicio de la transcripción (TSS). Estos descubrimientos indicaron que los transcritos nacientes son vulnerables a la escisión y poliadenilación prematuras (PCPA) y que U1 tiene una función crítica en la protección de los pre-ARNm frente a este proceso potencialmente destructor. Se mostró además que la supresión de PCPA es una función separada, independiente de corte y empalme y específica de U1, ya que no ocurría cuando el corte y empalme se inhibió con AMO de ARNsn de U2 o el inhibidor de corte y empalme, espliceostatina A (SSA) (Kaida et al., 2007, Nat Chem Biol 3:576-583).

Estas observaciones se hicieron por perfilado del transcriptoma usando matrices en mosaico de genoma parcial, lo que proporciona una información limitada. Además, en US2003/082149 y EP 1 384784 se mostró que un ARN mensajero diana o expresión génica, respectivamente, puede inhibirse con un ARN nuclear pequeño de U1 modificado. En WO 94/20141 se describió un anticuerpo monoclonal frente a la RNPsn U1. La presente invención aborda estos asuntos de información limitada y cumple con la necesidad en la técnica para evaluar exhaustivamente el papel de la RNPsn U1.

Resumen de la invención

La invención se refiere a un ácido nucleico que codifica un ARNsn de U1 endógeno, para uso en el tratamiento de una enfermedad o trastorno, en donde la enfermedad o trastorno se selecciona del grupo que consiste en cáncer, enfermedad autoinmune, inflamación, neurológica, y cualquier combinación de las mismas. Además, el ácido nucleico que codifica el ARNsn de U1 endógeno puede comprender además un plásmido de expresión de U1. La invención también se refiere a una molécula de ácido nucleico que incrementa el nivel de partículas de la RNPsn U1 endógena, para uso en el tratamiento de una enfermedad o trastorno, en donde la enfermedad o trastorno se selecciona del grupo que consiste en cáncer, enfermedad autoinmune, inflamación, neurológica, y cualquier combinación de las mismas y en donde la molécula de ácido nucleico es un plásmido que contiene el gen del ARNsn de U1 bajo el control de su promotor nativo.

También se describe un método para modular la expresión de un gen en una célula. En una realización, el método comprende poner en contacto una célula con un agente que regula uno o más de la actividad de la RNPsn U1 (U1) o nivel de U1, en donde uno o más de la actividad de U1 o nivel de U1 está asociado con la protección del transcrito naciente, modulando de esta manera la expresión de un gen en una célula.

En una realización, la actividad de U1 se selecciona del grupo que consiste en actividad de U1 asociada con la protección de un transcrito frente a la terminación prematura por escisión y poliadenilación (PCPA) en señales de poliadenilación crípticas (PAS), supresión por U1 de PCPA del transcrito naciente, actividad de U1 asociada con la protección de genes sin intrones frente a la escisión y poliadenilación prematuras, actividad de U1 asociada con el incremento de la longitud génica, actividad de U1 asociada con la transcripción más allá del extremo 3' canónico para generar nuevas secuencias de ARNm, actividad de U1 asociada con el transcriptoma en general, y cualquier combinación de los mismos.

En una realización, la supresión por U1 de PCPA del transcrito naciente se selecciona del grupo que consiste en intrones, exones, región no traducida en 5', región no traducida en 3', transcriptoma, y cualquier combinación de los mismos.

En una realización, una disminución en uno o más de la actividad de U1 y nivel de U1 da lugar a ARNm con regiones no traducidas en 3' más cortas.

En una realización, una disminución en uno o más de la actividad de U1 y nivel de U1 da lugar a isoformas con corte y empalme alternativo que se producen por el uso de sitios de poliadenilación alternativos (APA) más próximos.

En una realización, una disminución en uno o más de la actividad de U1 y nivel de U1 da lugar al acortamiento del transcrito eliminando de esta manera un elemento regulador del transcrito.

- 5 En una realización, el elemento regulador se selecciona del grupo que consiste en un dominio de unión de ARN, un dominio de unión de proteína, un sitio de unión de microARN, un sitio de unión de proteína RNPhn, y cualquier combinación de los mismos.

En una realización, el transcrito acortado da lugar a una expresión incrementada del gen.

- 10 En una realización, un incremento en uno o más de la actividad de U1 y nivel de U1 da lugar al alargamiento del transcrito evitando de esta manera la eliminación de un elemento regulador del transcrito.

En una realización, el elemento regulador se selecciona del grupo que consiste en un dominio de unión de ARN, un dominio de unión de proteína, un sitio de unión de microARN, un sitio de unión de proteína RNPhn, y cualquier combinación de los mismos.

En una realización, el transcrito alargado da lugar a una expresión disminuida del gen.

- 15 En una realización, el agente que regula uno o más de la actividad de U1 y nivel de U1 se selecciona del grupo que consiste en un agonista de U1, un antagonista de U1, y cualquier combinación de los mismos.

La descripción proporciona un método para modular el *telescripting* en una célula en donde el *telescripting* permite que los transcritos nacientes se extiendan determinando de esta manera uno o más de la longitud de ARNm y expresión de isoformas, comprendiendo el método regular uno o más de la actividad de la RNPsn U1 (U1) o nivel de U1 en la célula, en donde uno o más de la actividad de U1 o nivel de U1 está asociado con la protección del transcrito naciente.

- 20 La descripción proporciona un método para modular el *telescripting* en una célula en donde el *telescripting* permite que los transcritos nacientes se extiendan determinando de esta manera uno o más de la longitud de ARNm y expresión de isoformas, comprendiendo el método regular las cantidades relativas de los transcritos nacientes con uno o más de la actividad de U1 o nivel de U1 en la célula.

La descripción proporciona un método para modular el fenotipo celular de una célula, comprendiendo el método poner en contacto una célula con un agente que regula uno o más de la actividad de la RNPsn U1 (U1) o nivel de U1, en donde uno o más de la actividad de U1 o nivel de U1 está asociado con la protección del transcrito naciente, modulando de esta manera el fenotipo celular de la célula.

- 30 En una realización, el fenotipo celular se selecciona del grupo que consiste en proliferación, activación, diferenciación, oncogenicidad celular, y cualquier combinación de los mismos, activación de células inmunes, activación de neuronas, crecimiento de células madre, oncogenicidad, y cualquier combinación de los mismos.

- 35 En una realización, el fenotipo celular se selecciona del grupo que consiste en activación de células inmunes, activación de neuronas, crecimiento de células madre, oncogenicidad de una célula cancerosa, y cualquier combinación de los mismos.

En una realización, el incremento de uno o más de la actividad de U1 y nivel de U1 atenúa la oncogenicidad de la célula.

En una realización, el incremento de uno o más de la actividad de U1 y nivel de U1 inhibe el acortamiento de ARNm en neuronas activadas.

- 40 La descripción proporciona un método para tratar una enfermedad o trastorno en un sujeto. En una realización, el método comprende administrar un agente que regula uno o más de la actividad de la RNPsn U1 (U1) o nivel de U1, en donde uno o más de la actividad de U1 o nivel de U1 está asociado con la protección del transcrito naciente, tratando de esta manera la enfermedad o trastorno en el sujeto.

- 45 En una realización, la enfermedad o trastorno se selecciona del grupo que consiste en cáncer, enfermedad autoinmune, inflamación, infección, neurológica, y cualquier combinación de los mismos.

En una realización, el agente es un agonista de U1, además en donde el agonista de U1 da lugar al alargamiento del transcrito evitando de esta manera la eliminación de un elemento regulador del transcrito.

En una realización, el elemento regulador se selecciona del grupo que consiste en un dominio de unión de ARN, un dominio de unión de proteína, un sitio de unión de microARN, un sitio de unión de proteína RNPhn, y cualquier combinación de los mismos.

En una realización, el transcrito alargado da lugar a una expresión disminuida del gen. En una realización, el gen es un protooncogén.

Descripción breve de los dibujos

Para el propósito de ilustrar la invención, en los dibujos se representan determinadas realizaciones de la invención. La Figura 1, que comprende la Figura 1A a la Figura 1C, representa los resultados de experimentos que demuestran que HIDE-seq identifica cambios en el transcriptoma de todo el genoma. (Figura 1A) Se muestran bibliotecas no sustractivas y sustractivas preparadas en ambas direcciones (Control-U1; U1-Control) (gel de la izquierda) junto con posibles productos de PCR de supresión. Se muestran secuencias expresadas diferencialmente, así como secuencias comunes. Se realizó PCR anidada (gel de la derecha) con cebadores que contenían códigos de barras (líneas de entramado) y los adaptadores 454 A y B. (Figura 1B). Se representaron las lecturas muestreadas aleatoriamente con más del 90 % de cobertura y 90 % de identidad frente a genes afectados para HeLa, NIH/3T3 y S2 de sustracciones individuales o combinadas. (Figura 1C) Las lecturas de HIDE-seq están por encima [U1-Control (ARRIBA)] y debajo [Control-U1 (ABAJO)] de estructuras génicas en negro (bloque=exón, bloque fino =UTR, línea horizontal=intrón). Las lecturas con corte y empalme están conectadas por líneas discontinuas, mientras las lecturas inmediatamente adyacentes separadas por sitios *RsaI*, *HaeIII*, y *AluI* no lo están. Veces de cambio en Log2 de GTA de células HeLa tratadas con AMO de U1 o SSA normalizados respecto a los controles (paneles inferiores). Las flechas apuntan a sitios de PCPA inferidos.

La Figura 2, que comprende la Figura 2A a la Figura 2D, representa los resultados de experimentos que ilustran que los perfiles de transcriptoma muestran una función conservada de la RNPsn U1 en la supresión de PCPA. (Figura 2A) Se muestran los perfiles de HIDE-seq junto con secuencias de lecturas poli(A) (localización indicada por flechas sólidas) alineadas con el genoma para S2, NIH/3T3, y HeLa. La distancia (nt) es relativa al sitio de corte y empalme en 5' con el hexámero PAS usado probablemente y se muestran cadenas de A añadidas a HIDE-seq. Véanse también las Figuras 9-11. (Figura 2B) Ilustraciones de patrones observados (representados por encima de cada columna) después de AMO de U1 con el porcentaje de cada categoría reportado para *Drosophila* (15), ratón (15) y ser humano a tres dosis (15, 1,0, 0,25) en comparación con el número total de genes afectados entre paréntesis. Intrones=cajas sólidas; exones=cajas vacías. (Figura 2C) Se muestran ejemplos individuales de patrones descritos en la Figura 2B con los paréntesis indicando los resultados de más de una concentración. Las flechas sólidas indican lecturas que finalizan en colas de poli(A); las flechas abiertas son PCPA inferidas; los triángulos son PAS documentados. Véanse también las Figuras 12-13. (Figura 2D) Usando las lecturas con colas de poli(A) terminando en intrones (Figuras 9-12; Tabla 2), se representa la distribución de las distancias de PCPA respecto al TSS para *Drosophila* y ratón (parte superior), y ser humano a AMO de U1 alto y bajo (parte inferior).

La Figura 3, que comprende la Figura 3A y la Figura 3B, representa los resultados de experimentos que demuestran que la PCPA inducida por una reducción moderada de U1 regula la expresión de las isoformas cortas. (Figura 3A) El perfil de HIDE-seq para GABPB1 a AMO de U1 alto y bajo están por encima de los diagramas de las isoformas largas y cortas. Se realizó RT-qPCR de las isoformas cortas y largas en ADNc de células transfectadas con un rango de concentraciones de AMO de U1 usando un cebador directo común y diferentes cebadores inversos como se indica (flechas). Las cajas punteadas indican la región de transcrito ensayada. Las relaciones de corta a larga se reportan por encima de los histogramas. (Figura 3B) Se ensayó UBAP2L por RT-PCR para las isoformas corta (debido a intercambio de exones en 3') y larga (geles) y por RT-qPCR en ADNc de células tratadas con varios AMO. Triángulo=sitio de corte y empalme en 3' y AMO de PAS, flechas sólidas=lecturas poli(A); flechas abiertas=PCPA inferida.

La Figura 4, que comprende la Figura 4A a la Figura 4E, representa los resultados de experimentos que ilustran que se requiere U1 unida fuera del sitio de corte y empalme en 5' para suprimir PAS>1 kb de distancia. (Figura 4A) Se midió la distribución de la distancia en log₁₀ a la que ocurría PCPA en cada organismo desde el inicio de la cola poli(A) hacia 5' del sitio de corte y empalme en 5' o 3' del sitio de corte y empalme en 3'. Véanse también las Figuras 9-11 y la Tabla 2. (Figura 4B) Las flechas indican múltiples sitios de PCPA por gen. (Figura 4C) Las células HeLa se transfectaron con el minigen NR3C1 WT (carril 1) o uno que contenía una mutación del sitio de corte y empalme en 5' (carriles 2-12) y cantidades crecientes de U1 mutante que se empareja por bases con una de cuatro localizaciones a lo largo del pre-ARNm. Se detectó PCPA en el intrón por 3'RACE con la RT-PCR del exón 2 sirviendo como el control de carga. El porcentaje de supresión (relación PCPA a exón 2) se normalizó respecto al carril 2. (Figura 4D) Se transfectaron AMO control o AMO de U1 junto con el minigen NR3C1 WT (carriles 1 y 2) o NR3C1 en el que la secuencia PAS1 [385 nt] en el intrón se duplicó (PAS2) y se puso 1295 nt en 3' del sitio de corte y empalme en 5'. Se indican los minigenes en los que el sitio de corte y empalme en 5' (carriles 3 y 4), PAS1 (carriles 5 y 6) o ambos (carriles 7 y 8) estaban mutados. Se realizó 3'RACE como en la (Figura 4C). (Figura 4E) Se transfectaron AMO control,

de U1 o un AMO frente al sitio de corte y empalme en 5' del gen BASP1 endógeno como se indica. Se midió PCPA en el PAS 3,5 kb en 3' del exón 1 por 3' RACE. Se muestra la RT-PCR en secuencias intrónicas en 5' (IR1) y 3' (IR2) del PAS.

La Figura 5, que comprende la Figura 5A y la Figura 5B, representa los resultados de experimentos que ilustran que una reducción moderada de U1 recapitula el intercambio de isoformas después de la activación neuronal. Las neuronas PC12 de rata (Figura 5A) y MN-1 de ratón (Figura 5B) se estimularon con forskolina o forskolina/KCl durante 3 hrs o se transfectaron separadamente con AMO control o cantidades crecientes de AMO de U1. Se muestran RT-PCR que detectan las formas cortas y largas de Homer-1/Vesl-1 y Dab-1 con los histogramas que representan la relación de S/L. Las cajas grises punteadas indican la región de transcrito ensayada, las flechas muestran las localizaciones del cebador, y G6PDH es un control de carga.

La Figura 6, que comprende la Figura 6A y la Figura 6B, representa los resultados de experimentos que demuestran que la relación U1 a pre-ARNm determina el cambio de poliadenilación inmediato-temprano. (Figura 6A) Se activaron PC12 con forskolina/KCl o se trataron de forma simulada (DMSO) durante los tiempos indicados y se marcaron con ³H-uridina durante 30 min antes de la recogida. Las relaciones de los transcritos nacientes al ARN total se determinaron por centelleo y se normalizaron respecto a las cantidades de ARN total no marcado. Los controles de DMSO representan el promedio de 6 puntos temporales por experimento y los valores activados representan el promedio de dos experimentos independientes (punto temporal de 3 hr medido una vez). El ARNs de U1 se midió por RT-qPCR en los mismos puntos temporales. (Figura 6B) Las células PC12 se transfectaron toda la noche (24 hrs) para sobreexpresar U1 (o vector vacío) y se activaron con forskolina/KCl. Las isoformas de Homer-1 se ensayaron por RT-PCR 3 hrs después de la activación, como en la Figura 5A. Los histogramas representan la relación de las formas S/L de experimentos biológicos repetidos con el porcentaje de sobreexpresión de U1 reportado más adelante.

La Figura 7 representa un modelo de PCPA cotranscripcional y su supresión por U1. Los factores de escisión y poliadenilación (CPSF) se asocian con TEC de pol II y están así listos para para la escisión y poliadenilación del extremo 3', pero también pueden causar terminación prematura (PCPA) de PAS crípticos encontrados a lo largo de los pre-ARNm. U1 se recluta en los transcritos nacientes por múltiples interacciones, incluyendo emparejamiento de bases de su extremo 5' con secuencias de pre-ARNm cognadas (sitio de corte y empalme en 5' y otro) y pueden suprimir PCPA mediante la inhibición de CPSF asociados con pol II sobre un rango de <1 kb. La escasez de U1 incrementa la probabilidad de que los PAS distales permanezcan desprotegidos.

La Figura 8, que comprende la Figura 8A a la Figura 8F, representa los resultados de experimentos que ilustran la sensibilidad de HIDE-seq y la validación de AMO de U1 de Drosophila. (Figura 8A) Diagrama de Venn de superposición en la acumulación de intrones entre los datos de HIDE-seq y las matices en mosaico para los cromosomas 5, 7 y 16 humanos (véanse los métodos discutidos en otro lugar de la presente memoria). (Figura 8B) Validación por RT-PCR de la retención predicha de intrones para POLR1C en células HeLa tratadas con AMO frente a las RNPs principales (U1 y U2) y menores (U12 y U6atac) indicadas. (Figura 8C) (izquierda) La predicción por HIDE-seq de omisión de exones en RAN se validó por RT-PCR en ADNc obtenido de células HeLa tratadas con AMO control, de U1, U2, U12, y U6atac. Los cebadores se localizaron en los exones 3 y 5 (caja=exón, línea=intrón; carril 1=marcador de tamaño; los productos esperados se ilustran a la derecha y sus tamaños se listan a continuación de la escalera). (superior derecha) PAM es otro ejemplo de omisión de exones en células tratadas con AMO de U1 detectado por HIDE-seq. Se indican la posición y longitud del exón omitido. (inferior derecha) APOA1BP muestra que las posiciones del sitio de corte y empalme tanto en 5' como en 3' están alteradas en las células tratadas con AMO de U1. Las longitudes de las secuencias diferenciales (subrayadas) se comparan con la longitud total de las lecturas en la sustracción opuesta (flecha). (Figura 8D) La retención de intrones (flechas) para CG5168 y Gint3 predicha por HIDE-seq para AMO de U1 en Drosophila se validó por RT-PCR en ADNc de células S2 tratadas con AMO control, de U1, U2, U12 y U6atac. (Figura 8E) Los ejemplos de PCPA identificadas por HIDE-seq se validaron por 3' RACE en las mismas muestras tratadas con AMO en la (Figura 8B) con cebadores directos en el exón 3 de Chic y en el exón 1 de genes Ten-m. PCPA es la escisión y poliadenilación prematuras en un intrón, PA2/PA1 son sitios de poliadenilación en la 3' UTR, y * indica un producto de PCR no específico. Los tamaños esperados de los productos de PCPA se listan por encima de las lecturas intrónicas. (Figura 8F) Los productos de 3' RACE de la (Figura 8C) se clonaron por TOPO-TA, se secuenciaron por Sanger, y se alinearon con el genoma de Drosophila para verificar que cada una representaba PCPA en intrones (secuencia inferior), así como para determinar si cada uno concordaba con la secuencia de lecturas obtenidas por HIDE-seq (secuencia superior). Las secuencias (acortadas aquí) del producto de 797 pb para Chic y el producto de 387 pb de Ten-m para 3' RACE confirmaron los resultados de HIDE-seq. Se muestran los PAS y colas de poli(A), así como la secuencia del cebador oligo(dT)-KpnIHindIIIxBal.

La Figura 9 representa el alineamiento y atributos de PCPA en seres humanos (HeLa). Las secuencias genómicas 75 nt en 5' y 3' de donde se añadieron las colas de poli(A) (2 nucleótidos del medio) respecto a lecturas que contenían intrones se listan para 71 lecturas de PCPA definitivas en seres humanos. Los nombres de los genes y las coordenadas genómicas de PCPA se listan a la izquierda. Los posibles hexámeros de PAS usados se resaltan en las secuencias.

Se listan las longitudes génicas (Gen L), distancias de PCPA desde el sitio de inicio de la transcripción (TSS), desde el TSS más cercano (aTSS), el número de intrones (No. de Int.) en los que se produce PCPA y su longitud (Int L), la distancia entre PCPA y el sitio de corte y empalme en 5' (5'ss), o el sitio de corte y empalme en 3' (3'ss). Los cálculos de la mediana, media y desviación estándar para cada uno se muestran en la Tabla 2. También se listan la distancia de un EST respecto a PCPA, el identificador de las isoformas alternativas reportadas en las bases de datos UCSC, Aceview, o EST, y el o los tejidos en los que se encontraron.

La Figura 10 representa el alineamiento y atributos de PCPA en ratones (3T3). Las secuencias genómicas 75 nt en 5' y 3' de donde se añadieron las colas de poli(A) (2 nucleótidos del medio) respecto a lecturas que contenían intrones se listan para 70 lecturas de PCPA definitivas en ratones. Los posibles hexámeros de PAS usados se resaltan en las secuencias. Se listan las longitudes génicas (Gen L), distancias de PCPA desde el sitio de inicio de la transcripción (TSS), desde el TSS más cercano (aTSS), el número de intrones (No. de Int.) en los que se produce PCPA y su longitud (Int L), la distancia entre PCPA y el sitio de corte y empalme en 5', o el sitio de corte y empalme en 3'. Los cálculos de la mediana, media y desviación estándar para cada uno se muestran en la Tabla 2.

La Figura 11 representa el alineamiento y atributos de PCPA en *Drosophila* (S2). Las secuencias genómicas 75 nt en 5' y 3' de donde se añadieron las colas de poli(A) (2 nucleótidos del medio) respecto a lecturas que contenían intrones se listan para 59 lecturas de PCPA definitivas en *Drosophila*. Los posibles hexámeros de PAS usados se resaltan en las secuencias. Se listan las longitudes génicas (Gen L), distancias de PCPA desde el sitio de inicio de la transcripción (TSS), desde el TSS más cercano (aTSS), el número de intrones (No. de Int.) en los que se produce PCPA y su longitud (Int L), la distancia entre PCPA y el sitio de corte y empalme en 5', o el sitio de corte y empalme en 3'. Los cálculos de la mediana, media y desviación estándar para cada uno se muestran en la Tabla 2.

La Figura 12 representa el alineamiento y atributos de PCPA en seres humanos (HeLa: 1,0 y 0,25 nmoles de AMO de U1). Las secuencias genómicas 75 nt en 5' y 3' de donde se añadieron las colas de poli(A) (2 nucleótidos del medio) respecto a lecturas que contenían intrones se listan para 50 lecturas de PCPA definitivas en seres humanos a las dosis bajas combinadas. Los posibles hexámeros de PAS usados se resaltan en las secuencias. Se listan las longitudes génicas (Gen L), distancias de PCPA desde el sitio de inicio de la transcripción (TSS), el número de intrones (No. de Int.) en los que se produce PCPA, la clase de PCPA, isoformas más cortas alternativas que coinciden con PCPA de bases de datos y los tejidos en los que se reportaron. Las abreviaturas para los tejidos son como sigue: Adeno Cx: adenocarcinoma; AnaOligo: oligodendroglioma anaplásico; BM: médula ósea; BR: mama; B: cerebro; Cx: carcinoma; CE: cerebelo; CO: colon; E: ojo; ESC: células madre embrionarias; H: corazón; HYP: hipotálamo; K: riñón; L: pulmón; LCC: carcinoma de células grandes; NB: neuroblastoma; O: ovario; P: páncreas; PGCT: tumores de células germinales combinados; PL: placenta; PR: próstata; RB: retinoblastoma; S: estómago; T: testículos; U: útero; V: hígado. Los cálculos de la mediana, media y desviación estándar para cada uno se muestran en la Tabla 2.

La Figura 13, que comprende la Figura 13A y la Figura 13B, representa los resultados de experimentos que demuestran que el acortamiento de las 3'UTR se encuentra en las células tratadas con AMO de U1. (Figura 13A). Se muestran los perfiles de HIDE-seq de 3'UTR para genes en células HeLa tratadas con tres dosis de AMO de U1 que presentaban acortamiento. (Figura 13B) Ejemplos adicionales como en la (Figura 13A) que también mostraron acortamiento de 3'UTR en tándem en células T activadas.

La Figura 14 representa los resultados de experimentos que demuestran que la sobreexpresión de U1 incrementa la longitud de las 3'UTR globalmente. Las células transfectadas con vector vacío control o plásmidos de expresión de U1 (1 µg y 1,5 µg, 24 hr) se marcaron con 4-tiouridina para seleccionar los transcritos nacientes. Los mapas de RNA-seq de genes representativos muestran el último exón con cajas punteadas indicando el cambio hacia 3'UTR más largas. Se muestran los sitios diana de miARN que están implicados en el cáncer y se expresan en HeLa.

La Figura 15 representa los resultados de experimentos que demuestran que una disminución moderada de U1 causa un acortamiento de ARNm global de las 3'UTR. Las células transfectadas con AMO control o de U1 (0,25 nmoles, 8 hr) se marcaron con 4-tiouridina para seleccionar y secuenciar los transcritos nacientes. Los mapas de RNA-seq de genes representativos muestran el último exón con cajas punteadas indicando regiones de acortamiento en la 3'UTR. Se realizó 3'RACE usando PCR anidada para detectar los ARNm poliadenilados. Los sitios diana de miARN se listan como en la Fig. 14.

La Figura 16, que comprende la Figura 16A a la Figura 16E, representa los resultados de experimentos que demuestran que la sobreexpresión de U1 disminuye la oncogenicidad de las células HeLa *in vitro*. (Figura 16A) Las células HeLa se transfectaron con el control (vector vacío) o plásmidos de expresión de la ARNs U1. La proliferación celular se midió por el ensayo MTT cada 24 horas. (Figura 16B, Figura 16C) La migración y la actividad invasiva celulares se midieron después de 24 horas. (Figura 16D, Figura 16E) Las células migratorias e invasoras se tiñeron 24 horas después de la incubación y se visualizaron por microscopía de contraste de fase (aumento 10X).

La Figura 17, que comprende la Figura 17A a la Figura 17E, representa los resultados de experimentos que demuestran que una disminución moderada de U1 aumenta la oncogenicidad de las células HeLa *in vitro*. (Figura 17A) Las células HeLa se transfectaron con las cantidades indicadas de AMO de U1 o AMO control. La proliferación celular se midió por el ensayo de viabilidad celular luminiscente Cell Titer-Glo cada 24 horas. (Figura 17B, Figura 17C) La migración y la actividad invasiva celulares se midieron 24 horas después de la transfección. Se cultivaron $1,5 \times 10^5$ y $2,5 \times 10^5$ células para los ensayos de migración e invasión, respectivamente. (Figura 17D, Figura 17E) Las células migratorias e invasoras se tiñeron 24 horas después de la incubación y se visualizaron por microscopía de contraste de fase (aumento (10X)).

La Figura 18 representa los resultados de experimentos que demuestran que los cambios en los niveles de U1 afectan a la traducción global. Las células transfectadas con plásmido control o $1,5 \mu\text{g}$ de plásmido de U1 (24 horas), así como $0,25 \text{ nmoles}$ de AMO de U1 (izquierda) o de U2 (derecha) (8 horas) se marcaron metabólicamente con [^{35}S]metionina durante 1 hora. La radiactividad incorporada en las proteínas sintetizadas *de novo*, que se precipitaron con TCA, se contó por centelleo.

La Figura 19, que comprende la Figura 19A y la Figura 19B, representa los resultados de experimentos que demuestran que la RNPSn U1 protege a los genes sin intrones de la escisión y poliadenilación prematuras, y su nivel modula la longitud del ARNm de los genes sin intrones bidireccionalmente. Los mapas de RNA-seq de ejemplos representativos que muestran los genes sin intrones completos con cajas punteadas indicando un cambio hacia transcritos más largos por la sobreexpresión de U1 (Figura 19A) y transcritos más cortos por AMO de U1 (Figura 19B).

La Figura 20 representa los resultados de experimentos que demuestran que la sobreexpresión de U1 da lugar a la transcripción más allá del extremo 3' canónico. Las células HeLa transfectadas con vector control vacío o plásmidos de expresión de U1 ($1 \mu\text{g}$ y $1,5 \mu\text{g}$, 24 hr) y marcadas con 4-tiouridina para seleccionar los transcritos nacientes. Los mapas de RNA-seq de genes representativos que muestran el último exón con cajas punteadas indicando un cambio hacia una UTR más larga que el extremo 3' canónico anotado. También se muestra la validación por RT-PCR y 3'RACE para unos pocos genes. Las flechas con punta negra indican PAS anotados, las puntas de flecha indican PAS identificados de nuevas. Se proporciona una representación esquemática de la UTR debajo de cada gen, incluyen el extremo 3' canónico anotado, y la UTR extendida. Las flechas indican la posición de los cebadores.

La Figura 21 representa una lista de genes relacionados con el cáncer afectados por cambios en la actividad de *telescripting* de U1 en células HeLa. El tratamiento con AMO de U1 ($0,25 \text{ nmoles}$) da lugar a 45 oncogenes con 3'UTR acortadas en comparación con HeLa control, mientras las UTR ya acortadas de 85 y después 20 oncogenes adicionales se restaura con un nivel incrementado de expresión del ARNSn de U1.

La Figura 22 es un gráfico que representa los resultados de experimentos que demuestran que la transfección de concentraciones crecientes de construcciones que expresan U1 incrementa el nivel total de ARNSn de U1. Los niveles totales de U1 en células HeLa transfectadas se determinaron por RT-qPCR 24 horas después de la transfección. La entrada de ARN se normalizó respecto al ARNr 5S.

La Figura 23, que comprende la Figura 23A a la Figura 23C, representa un resumen y análisis de la identificación de la terminación de la UTR de genes con cambio en la longitud de las 3'UTR. (Figura 23A) Diagramas de Venn representan el número de genes afectados por el acortamiento frente al alargamiento en muestras de AMO de U1 (parte superior) y sobreexpresión de U1 (parte inferior). (Figura 23B) Los niveles de U1 afectan el cambio en la longitud del ARNm en genes altamente transcritos con 3'UTR más largas. Se muestra la longitud de las 3'UTR en afectados frente a fondo (todos los genes expresados en HeLa o un conjunto de datos seleccionado aleatoriamente con un número igual de genes) para genes acortados (parte superior) y alargados (parte inferior). Se observa una UTR significativamente más larga (ensayo de suma de rangos de Wilcoxon, valor $p < 2,2 \times 10^{-16}$, respectivamente) para los genes afectados en comparación con el conjunto de datos total o seleccionado aleatoriamente. (Figura 23C) Se evaluaron los niveles de expresión (valores RPKM) como en la Figura 23B. Se observa una expresión significativamente mayor (ensayo de suma de rangos de Wilcoxon, valor $p < 2,2 \times 10^{-16}$) para los genes afectados en comparación con el conjunto de datos total o seleccionado aleatoriamente.

La Figura 24, que comprende la Figura 24A a la Figura 24C, representa los resultados de experimentos que demuestran que las disminuciones moderadas en U2 no promueven el fenotipo oncogénico en HeLa. Se realizaron ensayos de migración (Figura 24A) invasión (Figura 24B) y (Figura 24C) imagenología por contraste de fase 24 hrs después de la transfección con AMO control y de U2 ($0,05\text{-}1 \text{ nmoles}$) como se describe en la Figura 17.

Descripción detallada de la invención

La invención se basa, en parte, en el descubrimiento de que la RNPSn U1 (U1) está implicada en un mecanismo de regulación de la expresión génica global. Por ejemplo, los resultados presentados en la presente memoria demuestran el mapeo exitoso de sitios de escisión y poliadenilación prematuras (PCPA) de U1 en organismos divergentes. Además, la presente invención se refiere al descubrimiento de que U1, además de su papel en el corte y empalme,

protege a los pre-ARNm frente a una terminación prematura drástica por escisión y poliadenilación en señales de poliadenilación crípticas (PAS). Este proceso se refiere en la presente memoria como *telescripting*, que asegura la integridad del transcriptoma y regula la longitud de los ARNm. En algunos casos, el *telescripting* contribuye a que los transcritos nacientes se extiendan sobre largas distancias. Por consiguiente, el *telescripting* proporciona un nuevo mecanismo de regulación de la expresión génica global.

U1 también puede regular la expresión génica tanto en genes que comprenden intrones como en genes sin intrones. En una realización, U1 protege a los genes sin intrones de la escisión y poliadenilación prematuras. En otra realización, U1 modula la longitud del ARNm de genes sin intrones bidireccionalmente. En otra realización más, U1 incrementa la longitud génica. En una realización, U1 permite la transcripción más allá del extremo 3' canónico para generar nuevas secuencias de ARNm.

La invención también se basa parcialmente en el descubrimiento de que la sobreexpresión de U1 alargaba numerosas 3'UTR en células cancerosas y, de esta manera, se demuestra que el *telescripting* es una diana para moderar la tumorigénesis. Esto se debe a que el acortamiento de las 3'UTR elimina frecuentemente los sitios de unión de microARN (miARN) que están enriquecidos en esta parte de los ARNm, lo que da lugar a la pérdida de la regulación por los miARN. Esto alivia típicamente la represión de la traducción mediada por los miARN que da lugar frecuentemente a un incremento fuerte en la cantidad de proteína producida a partir del ARNm acortado, incluyendo protooncogenes.

Por consiguiente, la invención proporciona composiciones y métodos para regular todos los aspectos de la expresión génica. En una realización, la invención proporciona composiciones y métodos para regular U1 como un medio para regular la expresión génica. En una realización, una disminución en U1 disminuye las 3'UTR. En una realización, una disminución en U1 promueve la expresión génica. En una realización, una disminución en U1 promueve la expresión génica debido a la eliminación de dianas reguladoras en la 3' UTR. Un ejemplo de una diana reguladora incluye los sitios de unión de microARN (miARN) y otros elementos reguladores del ARNm.

En otra realización, un incremento en U1 incrementa las 3'UTR. En otra realización, un incremento en U1 inhibe la expresión génica. En una realización, un incremento en U1 inhibe la expresión génica debido a la prevención de la eliminación de dianas reguladoras en la 3' UTR. Un ejemplo de una diana reguladora incluye los sitios de unión de microARN (miARN) y otros elementos reguladores del ARNm.

En una realización, un incremento en U1 atenúa la migración y la capacidad invasiva celulares de una célula cancerosa, mientras su disminución aumente de forma dependiente de la dosis la migración y la capacidad invasiva celulares. Esto se debe a que, en algunos casos, muchos de los cambios en la longitud de las 3'UTR recapitulan la desregulación de la diana de miARN causada por el cáncer en oncogenes.

En una realización, la supresión de PCPA puede modularse disminuyendo la U1 disponible en un gran rango sin estar comprometido el corte y empalme. Por lo tanto, la supresión de PCPA proporciona un mecanismo de regulación de la expresión génica nuevo y global a través de la modulación de los niveles de U1.

Por consiguiente, la invención proporciona un método para modular U1 para aplicaciones terapéuticas incluyendo, pero no están limitados a 1) disminuir los niveles de U1 para estimular una respuesta inmune (p. ej., adyuvante en vacunas) o reparación neuronal; 2) incrementar los niveles de U1 (p. ej., mediante activadores que son moléculas pequeñas o sobreexpresión de U1) para disminuir las tasas de transcripción de genes asociados con el cáncer, tratando, de esta manera, el cáncer.

En una realización, la invención proporciona un método para disminuir los niveles de U1 con, por ejemplo, un ácido nucleico antisentido para aumentar las características celulares oncogénicas, incluyendo migración y capacidad invasiva celulares *in vitro*. Sin pretender la vinculación a ninguna teoría particular, se cree que la disminución de los niveles de U1 promueve el acortamiento del ARNm, eliminando los sitios de unión de micro ARN en la región no traducida en 3', dando lugar, de esta manera, a la pérdida de la regulación por miARN, que, a su vez, desreprime la expresión de proteínas a partir de los ARNm acortados, que, de esta manera, produce más proteína, incluyendo proteínas oncogénicas, antiapoptóticas, y semejantes. En otra realización, los antagonistas de U1 pueden disminuir y/o inhibir U1 con respecto al corte y empalme y protección de los pre-ARNm frente a la terminación prematura drástica por escisión y poliadenilación en señales de poliadenilación crípticas.

En otra realización, la invención proporciona un método para incrementar los niveles o actividad de U1 con el fin de atenuar la oncogenicidad. Por consiguiente, la invención también proporciona agonistas de U1 que incrementan el nivel y/o actividad de U1 con respecto al corte y empalme y protección de los pre-ARNm frente a la terminación prematura drástica por escisión y poliadenilación en señales de poliadenilación crípticas. En el contexto del cáncer, un agonista de U1 puede incrementar los niveles de U1 con el fin de prevenir el acortamiento del ARNm, y de esta manera prevenir la eliminación de dianas reguladoras en la 3' UTR (p. ej., sitios de unión de micro ARN). Un resultado del alargamiento de la 3'UTR usando un agonista de U1 permite la regulación del gen mediante las dianas reguladoras

en la 3' UTR. En una realización, el alargamiento de la 3'UTR da lugar a la represión de la expresión de proteínas a partir de los ARNm alargados, que, de esta manera, producen menos proteína, incluyendo proteínas oncogénicas, antiapoptóticas, y semejantes.

Definiciones

- 5 A no ser que se defina de otra forma, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que el entendido comúnmente por un experto en la técnica a la que pertenece la invención. Aunque pueden usarse cualesquiera métodos y materiales similares o equivalentes a los descritos en la presente memoria en la práctica de y/o para el ensayo de la presente invención, en la presente memoria se describen los materiales y métodos preferidos. En la descripción y reivindicación de la presente invención, se usará la siguiente terminología según cómo se define, cuando se proporciona una definición.

También debe entenderse que la terminología usada en la presente memoria es solo para el propósito de describir realizaciones particulares, y no se pretende que sea limitante.

- 15 Los artículos "un" y "una" se usan en la presente memoria para hacer referencia a uno o a más de uno (es decir, al menos uno) del objeto gramatical del artículo. Como ejemplo, "un elemento" significa un elemento o más de un elemento.

"Aproximadamente", tal y como se usa en la presente memoria cuando se hace referencia a un valor mensurable tal como una cantidad, una duración temporal, y semejantes, se pretende que englobe variaciones de $\pm 20\%$ o, en algunos casos, $\pm 10\%$, o, en algunos casos, $\pm 5\%$, o, en algunos casos, $\pm 1\%$, o, en algunos casos, $\pm 0,1\%$ del valor especificado, ya que dichas variaciones son apropiadas para realizar los métodos descritos.

- 20 El término "agente" se usa en la presente memoria para indicar un compuesto químico, una mezcla de compuestos químicos, una macromolécula biológica (tal como un ácido nucleico, un anticuerpo, una proteína o parte de la misma, p. ej., un péptido), o un extracto preparado a partir de materiales biológicos tales como bacterias, plantas, hongos, o células o tejidos de animales (particularmente, mamíferos). La actividad de dichos agentes puede hacer que sean adecuados como un "agente terapéutico" que es una sustancia (o sustancias) biológicamente, fisiológicamente, o farmacológicamente activa que actúa localmente o sistémicamente en un sujeto. Los agentes pueden comprender, por ejemplo, fármacos, metabolitos, intermedios, cofactores, análogos del estado de transición, iones, metales, toxinas y polímeros naturales y sintéticos (p. ej., proteínas, péptidos, ácidos nucleicos, polisacáridos, glicoproteínas, hormonas, receptores y superficies celulares tales como paredes celulares y membranas celulares. Los agentes también pueden comprender alcoholes, haluros de alquilo, aminas, amidas, ésteres, aldehídos, éteres y otras clases de agentes orgánicos.

Tal y como se usa en la presente memoria, un "agonista de U1" se refiere a cualquier agente que mimetiza, activa, estimula, potencia o incrementa la actividad biológica de U1. Un agonista de U1 puede ser U1 o un fragmento de la misma; un agente que mimetiza U1 (tal como una molécula pequeña); un agente que incrementa o aumenta la expresión de U1; un agente que aumenta la unión de U1 a ácido nucleico o su compañero de unión; y semejantes.

- 35 Tal y como se usa en la presente memoria, un "antagonista de U1" se refiere a cualquier agente que atenúa, inhibe, se opone, contrarresta, o disminuye la actividad biológica de U1. Un antagonista de U1 puede ser un agente que inhibe o neutraliza la actividad de U1 (incluyendo, sin limitación, moléculas pequeñas y anticuerpos anti-U1); un agente que inhibe o disminuye la expresión de U1 (incluyendo, sin limitación, una molécula antisentido, una molécula de ARNi, y semejantes); un agente que previene la unión de U1 a ácido nucleico o un compañero de unión de U1.

- 40 El término "anticuerpo", tal y como se usa en la presente memoria, se refiere a una molécula de inmunoglobulina que es capaz de unirse específicamente a un epítipo específico en un antígeno. Los anticuerpos pueden ser inmunoglobulinas intactas derivadas de fuentes naturales o de fuentes recombinantes y pueden ser partes inmunorreactivas de inmunoglobulinas intactas. Los anticuerpos son típicamente tetrámeros de moléculas de inmunoglobulina. Los anticuerpos en la presente invención pueden existir en una variedad de formas incluyendo, por ejemplo, anticuerpos policlonales, anticuerpos monoclonales, Fv, Fab y F(ab)₂, así como anticuerpos de cadena única y anticuerpos humanizados (Harlow et al., 1999, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY; Harlow et al., 1989, Antibodies: A Laboratory Manual, Cold Spring Harbor, N.Y.; Houston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; Bird et al., 1988, Science 242:423-426).

- 50 "Antisentido" se refiere particularmente a la secuencia de ácido nucleico de la cadena no codificadora de una molécula de ADN bicatenario que codifica un polipéptido, o a una secuencia que es sustancialmente homóloga a la cadena no codificadora. Como se define en la presente memoria, una secuencia antisentido es complementaria a la secuencia de una molécula de ADN bicatenario que codifica un polipéptido. No es necesario que la secuencia antisentido sea complementaria solamente a la parte codificadora de la cadena codificadora de la molécula de ADN. La secuencia antisentido puede ser complementaria a las secuencias reguladoras especificadas en la cadena codificadora de una

molécula de ADN que codifica un polipéptido, secuencias reguladoras que controlan la expresión de las secuencias codificadoras.

El término “efecto antitumoral”, tal y como se usa en la presente memoria, se refiere a un efecto biológico que puede manifestarse por una disminución en el volumen tumoral, una disminución en el número de células tumorales, una

disminución en el número metástasis, un incremento en la esperanza de vida, o mejora de varios síntomas fisiológicos asociados con la afección cancerosa.

El término “cáncer”, tal y como se usa en la presente memoria, se define como una enfermedad caracterizada por el crecimiento rápido e incontrolado de células aberrantes. Las células cancerosas pueden diseminarse localmente o a través de la corriente sanguínea y el sistema linfático a otras partes del cuerpo. Los ejemplos de varios cánceres incluyen, pero no están limitados a, cáncer de mama, cáncer de próstata, cáncer de ovario, cáncer de cuello uterino, cáncer de piel, cáncer pancreático, cáncer colorrectal, cáncer renal, cáncer de hígado, cáncer de cerebro, linfoma, leucemia, cáncer de pulmón y semejantes.

“Que codifica” se refiere a la propiedad inherente de secuencias de nucleótidos específicas en un polinucleótido, tal como un gen, un ADNc, o un ARNm, para funcionar como moldes para la síntesis de otros polímeros y macromoléculas en procesos biológicos que tienen bien una secuencia de nucleótidos definida (p. ej., ARNr, ARNt y ARNm) o una secuencia de aminoácidos definida y las propiedades biológicas que resultan de ellas. Así, un gen codifica una proteína, si la transcripción y traducción del ARNm correspondiente a ese gen produce la proteína en una célula u otro sistema biológico. Tanto la cadena codificadora, cuya secuencia de nucleótidos es idéntica a la secuencia de ARNm y se proporciona habitualmente en listados de secuencias, como la cadena no codificadora, usada como el molde para la transcripción de un gen o ADNc, puede referirse como que codifica la proteína u otro producto de ese gen o ADNc.

“Cantidad efectiva” o “cantidad terapéuticamente efectiva” se usan indistintamente en la presente memoria, y se refieren a una cantidad de un compuesto, formulación, material, o composición, como se describe en la presente memoria, efectiva para conseguir un resultado biológico particular. Dichos resultados pueden incluir, pero no están limitados a, la inhibición de una infección viral como se determina por cualquier medio adecuado en la técnica.

El término “expresión”, tal y como se usa en la presente memoria, se define como la transcripción y/o traducción de una secuencia de nucleótidos particular dirigida por su promotor.

“Homólogo”, tal y como se usa en la presente memoria, se refiere a la similitud en la secuencia de subunidades entre dos moléculas poliméricas, p. ej., entre dos moléculas de ácido nucleico, p. ej., dos moléculas de ADN o dos moléculas de ARN, o entre dos moléculas de polipéptido. Cuando una posición en la subunidad en ambas de las dos moléculas está ocupada por la misma subunidad monomérica, p. ej., si una posición en cada una de dos moléculas de ADN está ocupada por adenina, entonces son homólogas en esa posición. La homología entre dos secuencias es una función directa del número de posiciones concordantes u homólogas, p. ej., si la mitad (p. ej., cinco posiciones en un polímero con una longitud de diez subunidades) de las posiciones en dos secuencias combinadas son homólogas, entonces las dos secuencias son homólogas en un 50 %, si el 90 % de las posiciones, p. ej., 9 de 10, concuerdan o son homólogas, las dos secuencias comparten un 90 % de homología.

Tal y como se usan en la presente memoria, los términos “gen” y “gen recombinante” se refieren a moléculas de ácido nucleico que comprenden un marco de lectura abierto que codifica un polipéptido de la invención. Dichas variaciones alélicas naturales pueden dar lugar típicamente a una varianza del 1-5 % en la secuencia de nucleótidos de un gen dado. Los alelos alternativos pueden identificarse mediante la secuenciación del gen de interés en varios individuos diferentes. Esto puede llevarse a cabo fácilmente usando sondas de hibridación para identificar el mismo locus genético en una variedad de individuos. Se pretende que cualquiera y todas de dichas variaciones de nucleótidos y polimorfismos o variaciones de aminoácidos resultantes que son el resultado de la variación alélica natural y que no alteran la actividad funcional estén dentro del alcance de la invención.

El término “inhibe”, tal y como se usa en la presente memoria, significa que suprime o bloquea una actividad o función al menos aproximadamente un diez por ciento respecto a un valor control. Preferiblemente, la actividad se suprime o bloquea un 50 % en comparación con un valor control, más preferiblemente un 75 %, e incluso más preferiblemente un 95 %.

Un “ácido nucleico aislado” se refiere a un segmento o fragmento de ácido nucleico que se ha separado de secuencias que lo flanquean en un estado natural, p. ej., un fragmento de ADN que se ha retirado de las secuencias que normalmente están adyacentes al fragmento, p. ej., las secuencias adyacentes al fragmento en un genoma en el que naturalmente aparece. El término también se aplica a ácidos nucleicos, que se han purificado sustancialmente de otros componentes, que naturalmente acompañan al ácido nucleico, p. ej., ARN o ADN o proteínas, que naturalmente le acompañan en la célula. El término por lo tanto incluye, por ejemplo, un ADN recombinante que se incorpora en un vector, en un plásmido que se replica de forma autónoma o virus, o en el ADN genómico de un procariota o eucariota, o que existe como una molécula separada (p. ej., como un ADNc o un fragmento genómico o de ADNc producido por

PCR o digestión con enzimas de restricción) independiente de otras secuencias. También incluye un ADN recombinante, que es parte de un gen híbrido que codifica una secuencia de polipéptido adicional.

- Tal y como se usa en la presente memoria, el término “modula” se pretende que se refiera a cualquier cambio en un estado biológico, es decir, incremento, disminución, y semejantes. Por ejemplo, el término “modula” se refiere a la capacidad de regular, positivamente o negativamente, la expresión o actividad de U1, incluyendo, pero no limitado a, la transcripción del ARNm de U1, estabilidad del ARNm de U1, traducción del ARNm de U1, estabilidad del polipéptido U1, modificaciones posteriores a la traducción de U1, o cualquier combinación de los mismos. Además, el término modula puede usarse para hacer referencia a un incremento, disminución, enmascaramiento, alteración, supresión o restauración de la actividad, incluyendo, pero no limitado a, la actividad de U1 asociada con su papel en el corte y empalme y papel en la protección de los pre-ARNm frente a la terminación prematura drástica por escisión y poliadenilación (PCPA) en señales de poliadenilación crípticas (PAS) en intrones, actividad de U1 asociada con la protección de genes sin intrones frente a la escisión y poliadenilación prematuras, actividad de U1 asociada con el incremento de la longitud génica, actividad de U1 asociada con la transcripción más allá del extremo 3' canónico para generar nuevas secuencias de ARNm, y actividad de U1 asociada con el transcriptoma en general.
- 15 Cuando se describe que dos polinucleótidos están “unidos operativamente” se quiere decir que un resto de ácido nucleico monocatenario o bicatenario comprende los dos polinucleótidos dispuestos en el resto de ácido nucleico de manera tal que al menos uno de los dos polinucleótidos es capaz de ejercer un efecto fisiológico por el cual se caracteriza sobre el otro. Como ejemplo, un promotor unido operativamente a la región codificadora de un gen es capaz de promover la transcripción de la región codificadora.
- 20 Preferiblemente, cuando el ácido nucleico que codifica la proteína deseada comprende además una secuencia promotora/reguladora, la secuencia promotora/reguladora está posicionada en el extremo 5' de la secuencia codificadora de la proteína deseada de manera que dirige la expresión de la proteína deseada en una célula. Conjuntamente, el ácido nucleico que codifica la proteína deseada y su secuencia promotora/reguladora comprenden un “transgén”.
- 25 La expresión “constitutiva” es un estado en el que un producto génico se produce en una célula viva bajo la mayoría o todas las condiciones fisiológicas de la célula.
- La expresión “inducible” es un estado en el que un producto génico se produce en una célula viva en respuesta a la presencia de una señal en la célula.
- Un “polipéptido recombinante” es uno que se produce después de la expresión de un polinucleótido recombinante.
- 30 “Polipéptido” se refiere a un polímero compuesto por residuos de aminoácidos, variantes estructurales naturales relacionadas, y análogos sintéticos no naturales de los mismos unidos mediante enlaces peptídicos, variantes estructurales naturales relacionadas, y análogos sintéticos no naturales de los mismos. Los polipéptidos sintéticos pueden sintetizarse, por ejemplo, usando un sintetizador de polipéptidos automatizado.
- El término “proteína” se refiere típicamente a polipéptidos largos.
- 35 El término “péptido” se refiere típicamente a polipéptidos cortos.

Tal y como se usa en la presente memoria, “tratar” significa reducir la frecuencia con la que un paciente experimenta los síntomas de la enfermedad, o alterar el historial y/o progresión natural de la enfermedad en un paciente.

Un tratamiento “terapéutico” es un tratamiento administrado a un sujeto que presenta signos de patología para el propósito de disminuir o eliminar estos signos.

- 40 Rangos: a lo largo de esta descripción, varios aspectos de la invención pueden presentarse en un formato de rango. Debe entenderse que la descripción en formato de rango es meramente por conveniencia y brevedad y no debe considerarse como una limitación inflexible en el alcance de la invención. Por consiguiente, debe considerarse que la descripción de un rango ha descrito específicamente todos los subrangos posibles, así como los valores numéricos individuales en ese rango. Por ejemplo, debe considerarse que la descripción de un rango tal como de 1 a 6 ha descrito específicamente los subrangos tales como de 1 a 3, de 1 a 4, de 1 a 5, de 2 a 4, de 2 a 6, de 3 a 6 etc., así como los números individuales en ese rango, por ejemplo, 1, 2, 2,7, 3, 4, 5, 5,3, y 6. Esto se aplica independientemente de la amplitud del rango.
- 45

Descripción

- 50 La invención proporciona composiciones y métodos para regular la expresión génica. En una realización, las composiciones y métodos de la invención están dirigidos a un proceso referido en la presente memoria como *telescripting*. En una realización, el *telescripting* asegura la integridad del transcriptoma y regula la longitud de los

ARNm. En algunos casos, el *telescripting* contribuye a que los transcritos nacientes se extiendan sobre grandes distancias. Por consiguiente, el *telescripting* proporciona un nuevo mecanismo de regulación de la expresión génica global.

5 En una realización, la invención proporciona composiciones y métodos para regular la actividad de U1 asociada con su papel en el corte y empalme, así como su papel en la protección de los pre-ARNm frente a terminación prematura drástica por escisión y poliadenilación (PCPA) en señales de poliadenilación críptica (PAS) en intrones.

10 En otra realización, la invención proporciona composiciones y métodos para regular U1, en donde U1 regula la expresión génica en genes que comprenden intrones o genes sin intrones. En una realización, U1 protege a los genes sin intrones de la escisión y poliadenilación prematuras. En otra realización, U1 modula la longitud del ARNm de genes sin intrones bidireccionalmente. En otra realización más, U1 incrementa la longitud génica. En una realización, U1 permite la transcripción más allá del extremo 3' canónico para generar nuevas secuencias de ARNm.

15 En una realización, la invención proporciona composiciones y métodos para regular todos los aspectos de la expresión génica. En una realización, la invención proporciona composiciones y métodos para regular U1 como un medio para regular la expresión génica. En una realización, un incremento en U1 alarga numerosas 3'UTR. En una realización, el alargamiento de las 3'UTR previene la eliminación de sitios de unión reguladores (p. ej., sitios de unión de miARN) enriquecidos en esa parte de los ARNm. En algunos casos, el alargamiento de las 3'UTR da lugar a una regulación incrementada por los miRNA, disminuyendo de esta manera la expresión. Por lo tanto, en algunos casos, un incremento en U1 inhibe la expresión génica.

20 En una realización, una disminución en U1 acorta numerosas 3'UTR. En una realización, el acortamiento de las 3'UTR elimina los sitios de unión reguladores (p. ej., sitios de unión de miARN) enriquecidos en esa parte de los ARNm. En algunos casos, el acortamiento de las 3'UTR da lugar a una regulación disminuida por los miARN, incrementando de esta manera la expresión. Por lo tanto, en algunos casos, una disminución en U1 incrementa la expresión génica.

25 En una realización, la invención proporciona un método para reducir la expresión de un gen en una célula que comprende incrementar U1 en la célula. En otra realización, la invención proporciona un método para reducir la producción de proteínas en una célula que comprende incrementar U1 en la célula. En otra realización, la invención proporciona un método para reducir el rendimiento en una célula de un ARNm producido por un gen en la célula que comprende incrementar U1 en la célula.

30 En una realización, la invención proporciona un método para incrementar la expresión de un gen en una célula que comprende disminuir U1 en la célula. En otra realización, la invención proporciona un método para incrementar la producción de proteínas en una célula que comprende disminuir U1 en la célula. En otra realización, la invención proporciona un método para incrementar el rendimiento en una célula de un ARNm producido por un gen en la célula que comprende incrementar U1 en la célula.

35 En una realización, la disminución de los niveles y/o actividad de U1 da lugar a un aumento en las características oncogénicas de una célula, incluyendo, pero no limitado a, migración y capacidad invasiva celulares. En otra realización, el incremento o sobreexpresión de U1 atenúa la oncogenicidad. Sin pretender la vinculación a ninguna teoría particular, se cree que una disminución en los niveles y/o actividad de U1 promueve el acortamiento del ARNm, lo que elimina sitios de unión de miARN en la región no traducida en 3' de una molécula de pre-ARNm, desreprimiendo de esta manera la expresión de proteínas a partir de los ARNm acortados y, por lo tanto, produciendo más proteína, incluyendo proteínas oncogénicas, antiapoptóticas, y semejantes.

40 Por consiguiente, la invención proporciona composiciones y métodos para tomar como diana y regular U1 y sus equivalentes funcionales en una célula para regular la expresión de proteínas. En una realización, la regulación de U1 y sus equivalentes funcionales en una célula permite la regulación de una característica celular. En algunos casos, la característica celular que se está regulando a través de la regulación de U1 es la característica oncogénica de la célula. Por ejemplo, el incremento o activación de U1 en una célula da lugar a una atenuación de la oncogenicidad de una célula. Alternativamente, la disminución o inhibición de U1 en una célula da lugar a un incremento en la oncogenicidad de una célula.

45 La invención no debe limitarse solo a regular la oncogenicidad de una célula. En lugar de esto, la invención permite la regulación de cualquier fenotipo de una célula en donde el fenotipo está asociado con la expresión de una proteína o variante de corte y empalme particular. Esto es, la regulación de U1 permite la regulación de la expresión de una proteína o variante de corte y empalme deseada. Por ejemplo, cuando se desea incrementar la expresión de una proteína particular, U1 puede inhibirse, de manera que el pre-ARNm asociado con la proteína deseada se acorta hasta el grado en el que se eliminan los sitios de unión de micro ARN en la región no traducida en 3', desreprimiendo de esta manera la expresión a partir de los ARNm acortados que, a su vez, producen más proteína. Alternativamente, cuando se desea disminuir la expresión de una proteína particular, U1 puede activarse o incrementarse, de manera que el pre-ARNm asociado con la proteína deseada no se acorta, de manera que están presentes los sitios de unión

de micro ARN en la región no traducida en 3', reprimiendo de esta manera la expresión a partir de los ARNm, que, a su vez, producen menos proteína.

Composiciones

5 La presente invención comprende cualquier agente que mimetiza, activa, estimula, potencia o incrementa la actividad biológica de la RNPs_n U1 (U1) que puede usarse como un agonista de U1. La invención también comprende cualquier agente que inhibe, se opone, contrarresta, o disminuye la actividad biológica de U1 que puede usarse como un antagonista de U1.

10 Los agonistas de U1 de la invención incluyen, pero no están limitados a, ácidos nucleicos purificados o recombinantes que codifican proteínas U1 o fragmentos de las mismas; polipéptidos U1 purificados o recombinantes o fragmentos de los mismos; u otros agentes que mimetizan, activan, estimulan, potencian o incrementan la actividad biológica de U1. Los ejemplos de agonista de U1 incluyen, sin limitación, agentes que incrementan el ARNm de U1 o la expresión de la proteína; agentes que incrementan la interacción entre U1 y ácido nucleico, y otros agentes que se identifican por cualquiera de los métodos de cribado descritos en la presente memoria o en el futuro.

15 En una realización, la invención comprende una composición para atenuar la oncogenicidad. Por ejemplo, un agonista de U1 de la invención puede usarse para incrementar o sobreexpresar U1, en donde U1 es capaz de prevenir el acortamiento del pre-ARNm y/o promover el alargamiento del pre-ARNm. En algunos casos, un agonista de U1 de la invención puede usarse para incrementar o sobreexpresar U1, en donde U1 es capaz de proteger a los pre-ARNm de la terminación prematura drástica por escisión y poliadenilación en señales de poliadenilación críptica y evita de esta manera la eliminación de sitios de unión de micro ARN en la región no traducida en 3', que puede dar lugar a la represión de la expresión de proteínas de una proteína oncogénica. En otra realización, el agonista de U1 puede regular tanto genes con intrones como sin intrones.

20 En una realización, la invención comprende una composición para inhibir U1 (p. ej., antagonistas de U1). Por ejemplo, un antagonista de U1 de la invención puede usarse para inactivar o disminuir la expresión de U1, en donde la inactivación o disminución de la expresión de U1 acorta la longitud del pre-ARNm. En algunos casos, el acortamiento del pre-ARNm elimina sitios de unión reguladores (p. ej., miARN) en el pre-ARNm, lo que puede dar lugar a una expresión incrementada. En una realización, el acortamiento del pre-ARNm elimina sitios de unión reguladores (p. ej., miARN) en el pre-ARNm, lo que puede dar lugar a la desrepresión de la expresión de proteínas de una proteína oncogénica. En otra realización, el antagonista de U1 puede regular tanto genes con intrones como sin intrones.

30 Los antagonistas de U1 de la invención incluyen, pero no están limitados a, agentes que inactivan, inhiben, sequestran o disminuyen la actividad biológica de U1. El ejemplo de antagonista de U1 incluyen, sin limitación, un ARN de interferencia pequeño (ARNsi), un microARN, un ácido nucleico antisentido, una ribozima, un vector de expresión que codifica un mutante negativo transdominante, un anticuerpo intracelular, un péptido y una molécula pequeña.

35 Un polinucleótido ARNsi es una molécula de ácido nucleico ARN que interfiere con la actividad del ARN que se considera que ocurre generalmente a través de un mecanismo de silenciamiento génico posterior a la transcripción. Un polinucleótido ARNsi comprende preferiblemente un ARN bicatenario (ARNds) pero no se pretende que esté así limitado y puede comprender un ARN monocatenario (véase, p. ej., Martínez et al., 2002 Cell 110:563-74). El polinucleótido ARNsi incluido en la invención puede comprender otros polímeros de nucleótidos monocatenarios o bicatenarios naturales, recombinantes, o sintéticos (ribonucleótidos o desoxirribonucleótidos o una combinación de ambos) y/o análogos de nucleótidos como se proporciona en la presente memoria (p. ej., un oligonucleótido o polinucleótido o semejantes, típicamente en unión fosfodiéster 5' a 3'). Por consiguiente, se apreciará que también se pretende que determinadas secuencias ejemplares descritas en la presente memoria como secuencias de ADN capaces de dirigir la transcripción de los polinucleótidos ARNsi describan las secuencias de ARN correspondientes y sus complementos, dados los principios bien establecidos del emparejamiento de bases complementarias de nucleótidos.

45 En otra realización, la secuencia moduladora es una molécula de ácido nucleico antisentido. En algunos casos, la secuencia de ácido nucleico antisentido puede expresarse por un vector. Por ejemplo, un vector de expresión antisentido puede usarse para transfectar una célula de mamífero o al mamífero en sí mismo, causando de esta manera una expresión endógena reducida de un regulador deseado (p. ej., U1) en la célula. Sin embargo, no debe considerarse que la invención está limitada a inhibir la expresión de un regulador por transfección de células con moléculas antisentido. En lugar de esto, la invención engloba otros métodos conocidos en la técnica para inhibir la expresión o actividad de una proteína en la célula incluyendo, pero no limitado a, el uso de una ribozima, la expresión de un regulador no funcional (es decir, mutante negativo transdominante) y el uso de un anticuerpo intracelular.

50 Las moléculas antisentido y su uso para inhibir la expresión génica son muy conocidos en la técnica (véase, p. ej., Cohen, 1989, En: Oligodeoxyribonucleotides, Antisense Inhibitors of Gene Expression, CRC Press). Los ácidos nucleicos antisentido son moléculas de ADN o ARN que son complementarias, como se define este término en otro

lugar en la presente memoria, al menos a una parte de una molécula de ARNm específica (Weintraub, 1990, Scientific American 262:40). En la célula, los ácidos nucleicos antisentido hibridan con el ARNm correspondiente, formando una molécula bicatenaria, inhibiendo de esta manera la traducción de genes.

5 El uso de métodos antisentido para inhibir la traducción de genes es conocido en la técnica, y se describe, por ejemplo, en Marcus-Sakura (1988, Anal. Biochem. 172:289). Dichas moléculas antisentido pueden proporcionarse a la célula mediante expresión genética usando ADN que codifica la molécula antisentido como se enseña en Inoue, 1993, Pat. de EE. UU. No. 5.190.931.

10 Alternativamente, las moléculas antisentido de la invención pueden prepararse sintéticamente y proporcionarse entonces a la célula. Se prefieren los oligómeros antisentido de entre aproximadamente 10 a aproximadamente 30, y más preferiblemente aproximadamente 15 nucleótidos, ya que se sintetizan e introducen en una célula diana más fácilmente. Las moléculas antisentido sintéticas contempladas por la invención incluyen derivados de oligonucleótidos conocidos en la técnica que tienen una actividad biológica mejorada en comparación con los oligonucleótidos no modificados (véase la Pat. de EE. UU. No. 5.023.243).

15 Las ribozimas y su uso para inhibir la expresión génica también son muy conocidos en la técnica (véase, p. ej., Cech et al., 1992, J. Biol. Chem. 267:17479-17482; Hampel et al., 1989, Biochemistry 28:4929-4933; Eckstein et al., Publicación Internacional No. WO 92/07065; Altman et al., Pat. de EE. UU. No. 5.168.053). Las ribozimas son moléculas de ARN que poseen la capacidad de escindir específicamente otro ARN monocatenario de una manera análoga a las endonucleasas de restricción del ADN. A través de la modificación de secuencias de nucleótidos que codifican estos ARN, pueden prepararse por ingeniería moléculas que reconozcan secuencias de nucleótidos
20 específicas en una molécula de ARN y escindir las (Cech, 1988, J. Amer. Med. Assn. 260:3030). Una ventaja importante de esta estrategia es el hecho de que las ribozimas son específicas de secuencia.

25 Cualquier polinucleótido de la invención puede modificarse adicionalmente para incrementar su estabilidad *in vivo*. Las posibles modificaciones incluyen, pero no están limitadas a, la adición de secuencias flanqueantes en los extremos 5' y/o 3'; el uso de fósforotioato o 2' O-metilo en lugar de uniones fosfodiéster en el núcleo; y/o la inclusión de bases no tradicionales, tales como inosina, queosina, y wibutosina y semejantes, así como formas modificadas con acetilo, metilo, tio y otros de adenina, citidina, guanina, timina, y uridina.

30 En otro aspecto de la invención, U1 puede inhibirse mediante la inactivación y/o secuestro de U1. Como tal, la inhibición de los efectos de U1 puede conseguirse usando un mutante negativo transdominante. Alternativamente, puede usarse un anticuerpo específico de U1. En una realización, el antagonista es una proteína y/o compuesto que tiene la propiedad deseable de interactuar con un compañero de unión de U1 y, de esta manera, competir con U1. En otra realización, el antagonista es una proteína y/o compuesto que tiene la propiedad deseable de interactuar con U1 y, de esta manera, secuestrar a U1. Como entenderá un experto en la técnica, cualquier anticuerpo que pueda reconocer y unirse a un epítipo de interés (p. ej., un epítipo de U1) es útil en la presente invención.

35 La invención también incluye equivalentes funcionales de los anticuerpos descritos en la presente memoria. Los equivalentes funcionales tienen características de unión comparables a las de los anticuerpos, e incluyen, por ejemplo, anticuerpos hibridados y de cadena única, así como fragmentos de los mismos. Los métodos para producir dichos equivalentes funcionales se describen en la Solicitud PCT WO 93/21319 y Solicitud PCT WO 89/09622.

40 Los equivalentes funcionales incluyen polipéptidos con secuencias de aminoácidos sustancialmente iguales a la secuencia de aminoácidos de las regiones variables o hipervariables de los anticuerpos. La secuencia de aminoácidos "sustancialmente igual" se define en la presente memoria como una secuencia con al menos el 70 %, preferiblemente al menos aproximadamente el 80 %, más preferiblemente al menos aproximadamente el 90 %, incluso más preferiblemente al menos aproximadamente el 95 %, y lo más preferiblemente al menos el 99 % de homología con otra secuencia de aminoácidos (o cualquier número entero entre 70 y 99), como se determina por el método de búsqueda FASTA según Pearson y Lipman, 1988 *Proc. Nat'l. Acad. Sci. USA* 85: 2444-2448. Los anticuerpos
45 quiméricos u otros híbridos tienen regiones constantes derivadas sustancialmente o exclusivamente de regiones constantes de anticuerpos humanos y regiones variables derivadas sustancialmente o exclusivamente de la secuencia de la región variable de un anticuerpo monoclonal de cada hibridoma estable.

50 Los anticuerpos de cadena única (scFv) o fragmentos Fv son polipéptidos que consisten en la región variable de la cadena pesada del anticuerpo unida a la región variable de la cadena ligera, con o sin un conector de interconexión. Así, el Fv comprende un sitio de combinación de anticuerpo.

Modificación genética

La invención engloba vectores de expresión y métodos para la introducción de ADN exógeno en células con la expresión concomitante del ADN exógeno en las células, tales como los descritos, por ejemplo, en Sambrook et al.

(2012, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Nueva York), y en Ausubel et al. (1997, *Current Protocols in Molecular Biology*, John Wiley and Sons, Nueva York).

El polinucleótido deseado puede clonarse en varios tipos de vectores. Sin embargo, no debe considerarse que la presente invención está limitada a ningún vector particular. En lugar de esto, debe considerarse que la presente invención engloba una amplia plétora de vectores que están fácilmente disponibles y/o son muy conocidos en la técnica. Por ejemplo, un polinucleótido deseado de la invención puede clonarse en un vector incluyendo, pero no limitado a, un plásmido, un fagémido, un derivado de fago, un virus animal, y un cósmido. Los vectores con un interés particular incluyen vectores de expresión, vectores de replicación, vectores de generación de sondas, y vectores de secuenciación.

Además, el vector de expresión puede proporcionarse a una célula en la forma de un vector viral. La tecnología de vectores virales es muy conocida en la técnica y se describe, por ejemplo, en Sambrook et al. (2012), y en Ausubel et al. (1997), y en otros manuales de virología y biología molecular. Los virus, que son útiles como vectores incluyen, pero no están limitados a, retrovirus, adenovirus, virus adeno-asociados, virus de herpes, y lentivirus. En general, un vector adecuado contiene un origen de replicación funcional en al menos un organismo, una secuencia promotora, sitios de endonucleasa de restricción convenientes, y uno o más marcadores seleccionables. (Véase, p. ej., WO 01/96584; WO 01/29058; y Pat. de EE. UU. No. 6.326.193.

Para la expresión del polinucleótido deseado, al menos un módulo en cada promotor funciona para posicionar el sitio de inicio para la síntesis de ARN. El ejemplo mejor conocido de esto es la caja TATA, pero en algunos promotores que carecen de una caja TATA, tales como el promotor para el gen de la desoxinucleotidil transferasa terminal de mamíferos y el promotor para los genes de SV40, un elemento discreto superpuesto al sitio de inicio en sí mismo ayuda a fijar el lugar de inicio.

Los elementos promotores adicionales, es decir, potenciadores, regulan la frecuencia del inicio transcripcional. Típicamente, estos están localizados en la región 30-110 pb en 5' del sitio de inicio, aunque se ha mostrado que varios promotores también contienen elementos funcionales en 3' del sitio de inicio. El espaciamiento entre los elementos promotores es flexible frecuentemente, de manera que la función promotora se conserva cuando los elementos se invierten o se mueven uno respecto al otro. En el promotor de la timidina quinasa (tk), el espaciamiento entre los elementos promotores puede incrementarse hasta 50 pb antes de que la actividad empiece a disminuir. Dependiendo del promotor, parece que los elementos individuales pueden funcionar bien cooperativamente o independientemente para activar la transcripción.

Naturalmente, será importante emplear un promotor y/o potenciador que dirija efectivamente la expresión del segmento de ADN en el tipo de célula, orgánulo, y organismo elegido para la expresión. Los expertos en la técnica de la biología molecular saben generalmente cómo usar las combinaciones de promotores, potenciadores, y tipos celulares para la expresión de proteínas, por ejemplo, véase Sambrook et al. (2012). Los promotores empleados pueden ser constitutivos, específicos de tejido, inducibles, y/o útiles bajo las condiciones apropiadas para dirigir la expresión a alto nivel del segmento de ADN introducido, de manera que es ventajoso en la producción a gran escala de proteínas y/o péptidos recombinantes. El promotor puede ser heterólogo o endógeno.

El vector de expresión que se va a introducir en una célula también puede contener bien un gen de marcador seleccionable o un gen informador o ambos para facilitar la identificación y selección de las células que expresan a partir de la población de células que se desea transfectar o infectar mediante vectores virales. En otras realizaciones, el marcador seleccionable puede estar portado por una parte separada de ADN y usarse en un procedimiento de cotransfección. Tanto los genes de marcadores seleccionables como informadores pueden estar flanqueados por secuencias reguladoras apropiadas para permitir la expresión en las células huésped. Los marcadores seleccionables útiles se conocen en la técnica e incluyen, por ejemplo, genes de resistencia a antibióticos, tales como neo y semejantes.

Los genes informadores se usan para identificar células potencialmente transfectadas y para evaluar la funcionalidad de las secuencias reguladoras. Los genes informadores que codifican proteínas que se pueden ensayar fácilmente son muy conocidos en la técnica. En general, un gen informador es un gen que no está presente en o no se expresa por el organismo o tejido receptor y que codifica una proteína cuya expresión se manifiesta por alguna propiedad fácilmente detectable, p. ej., actividad enzimática. La expresión del gen informador se ensaya a un tiempo adecuado después de que el ADN se haya introducido en las células receptoras.

Los genes informadores adecuados pueden incluir genes que codifican luciferasa, beta-galactosidasa, cloranfenicol acetil transferasa, fosfatasa alcalina secretada, o el gen de la proteína fluorescente verde (véase, p. ej., Ui-Tei et al., 2000 FEBS Lett. 479:79-82). Los sistemas de expresión adecuados son muy conocidos y pueden prepararse usando técnicas muy conocidas u obtenerse comercialmente. Las construcciones de delección interna pueden generarse usando sitios de restricción internos únicos o por digestión parcial de sitios de restricción no únicos. Las construcciones

pueden transfectarse entonces en células que presentan altos niveles de expresión de polinucleótido ARNsi y/o polipéptido. En general, la construcción con la mínima región 5' flanqueante que muestra el mayor nivel de expresión del gen informador se identifica como el promotor. Dichas regiones promotoras pueden estar unidas a un gen informador y usarse para evaluar agentes para determinar su capacidad de modular la transcripción dirigida por el promotor.

En el contexto de un vector de expresión, el vector puede introducirse fácilmente en una célula huésped, p. ej., célula de mamífero, bacteriana, de levadura o insecto por cualquier método en la técnica. Por ejemplo, el vector de expresión puede transferirse a una célula huésped por medios físicos, químicos o biológicos. Se entiende fácilmente que la introducción del vector de expresión que comprende el polinucleótido de la invención rinde una célula silenciada con respecto a un regulador.

Los métodos físicos para introducir un polinucleótido en una célula huésped incluyen precipitación con fosfato de calcio, lipofección, bombardeo de partículas, microinyección, electroporación, y semejantes. Los métodos para producir células que comprenden vectores y/o ácidos nucleicos exógenos son muy conocidos en la técnica. Véase, por ejemplo, Sambrook et al. (2012, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Nueva York), y en Ausubel et al. (1997, *Current Protocols in Molecular Biology*, John Wiley and Sons, Nueva York).

Los métodos biológicos para introducir un polinucleótido de interés en una célula huésped incluyen el uso de vectores de ADN y ARN. Los vectores virales, y especialmente los vectores retrovirales, se han convertido en el método más ampliamente usado para insertar genes en células de mamíferos, p. ej., humanas. Otros vectores virales pueden derivar de lentivirus, virus de la viruela, virus del herpes simple I, adenovirus y virus adeno-asociados, y semejantes. Véase, por ejemplo, las Pat. de EE. UU. Nos. 5.350.674 y 5.585.362.

Los medios químicos para introducir un polinucleótido en una célula huésped incluyen sistemas de dispersión coloidal, tales como complejos de macromoléculas, nanocápsulas, microesferas, lechos, y sistemas basados en lípidos incluyendo emulsiones de aceite en agua, micelas, micelas mixtas, y liposomas. Un sistema coloidal preferido para uso como un vehículo de administración *in vitro* e *in vivo* es un liposoma (es decir, una vesícula de membrana artificial). La preparación y uso de dichos sistemas es muy conocido en la técnica.

Cualquier vector de ADN o vehículo de administración puede utilizarse para transferir el polinucleótido deseado en una célula *in vitro* o *in vivo*. En el caso en el que se utilice un sistema de administración no viral, un vehículo de administración preferido es un liposoma. Los sistemas de administración y protocolos mencionados anteriormente pueden encontrarse, por lo tanto, en *Gene Targeting Protocols*, 2ed., p 1-35 (2002) y *Gene Transfer and Expression Protocols*, Vol. 7, Murray ed., p 81-89 (1991).

“Liposoma” es un término genérico que engloba una variedad de vesículas lipídicas únicas y multilamelares formadas por la generación de bicapas o agregados lipídicos contenidos. Los liposomas pueden caracterizarse por tener estructuras vesiculares con una membrana de bicapa fosfolipídica y un medio acuoso interno. Los liposomas multilamelares tienen múltiples capas lipídicas separadas por medio acuoso. Se forman espontáneamente cuando los fosfolípidos se suspenden en un exceso de disolución acuosa. Los componentes lipídicos experimentan una autoreorganización antes de la formación de estructuras cerradas y atrapan agua y solutos disueltos entre las bicapas lipídicas. Sin embargo, la presente invención también engloba composiciones que tienen diferentes estructuras en disolución que la estructura vesicular normal. Por ejemplo, los lípidos pueden asumir una estructura micelar o existir meramente como agregados no uniformes de moléculas lipídicas. También se contemplan complejos de lipofectamina-ácido nucleico.

Métodos para modular la actividad de U1

La actividad de U1 puede modularse usando cualquier método descrito en la presente memoria o conocido por el experto en la técnica. Los ejemplos de métodos para aumentar la actividad de U1 incluyen, pero no están limitados a, el incremento de la expresión de un gen de U1 endógeno, incremento de la expresión de ARNm de U1, y aumento de la función, actividad, o estabilidad de una proteína U1. Un activador de U1 puede ser, por lo tanto, un compuesto que incrementa la expresión de un gen de U1, incrementa la semivida estabilidad y/o expresión del ARNm de U1, o aumenta la función, actividad, o estabilidad de la proteína U1. Un activador de U1 puede ser cualquier tipo de compuesto, incluyendo, pero no limitado a, un polipéptido, un ácido nucleico, un aptámero, un peptidomimético, y una molécula pequeña, o combinaciones de los mismos.

Los ejemplos de métodos para inhibir la actividad de U1 incluyen, pero no están limitados a, inhibir la expresión de un gen de U1 endógeno, disminuir la expresión de ARNm de U1, e inhibir la función, actividad, o estabilidad de una proteína U1. Un inhibidor de U1 puede ser, por lo tanto, un compuesto que disminuye la expresión de un gen de U1, disminuye la semivida estabilidad y/o expresión del ARNm de U1, o inhibe la función, actividad o estabilidad de la proteína U1. Un inhibidor de U1 puede ser cualquier tipo de compuesto, incluyendo, pero no limitado a, un polipéptido, un ácido nucleico, un aptámero, un peptidomimético, y una molécula pequeña, o combinaciones de los mismos.

La regulación de U1 puede conseguirse bien directamente o indirectamente. Por ejemplo, U1 puede inhibirse directamente por compuestos o composiciones que interactúan directamente con la proteína U1, tales como anticuerpos. Alternativamente, U1 puede inhibirse indirectamente por compuestos o composiciones que inhiben los efectores de U1 aguas abajo, o reguladores aguas arriba que regulan al alza la expresión de U1.

- 5 La disminución de la expresión de un gen de U1 endógeno incluye proporcionar un inhibidor específico de la expresión del gen de U1. La disminución de la expresión del ARNm de U1 o proteína U1 incluye disminuir la semivida o estabilidad del ARNm de U1 o disminuir la expresión del ARNm de U1. Los métodos para disminuir la expresión de U1 incluyen, pero no están limitados a, métodos que usan un ARNsi, un microARN, un ácido nucleico antisentido, una ribozima, un vector de expresión que codifica un mutante negativo transdominante, un péptido, una molécula pequeña, otros inhibidores específicos de la expresión del gen, ARNm, y proteína U1, y combinaciones de los mismos.

- 10 La invención incluye métodos para el tratamiento de un trastorno relacionado con U1. La administración de un inhibidor de U1 que comprende uno o más péptidos, una molécula pequeña, un ácido nucleico antisentido, un receptor soluble, un anticuerpo, o un agente que atenúa, inhibe, se opone, contrarresta, o disminuye la actividad biológica de U1 en un método de tratamiento puede conseguirse de varias maneras diferentes, usando métodos conocidos en la técnica. De manera similar, la administración de un activador de U1 que comprende uno o más péptidos, una molécula pequeña, o un agente que mimetiza, activa, estimula, potencia o incrementa la actividad biológica de U1 en un método de tratamiento puede conseguirse de varias maneras diferentes, usando métodos conocidos en la técnica.

- 15 En una realización, la invención incluye modular U1 para aplicaciones oncológicas. Por ejemplo, una disminución en los niveles de U1 da lugar al acortamiento de ARNm, incremento en la transcripción y cáncer (p. ej., acelera la proliferación, migración, capacidad invasiva). Por consiguiente, el tratamiento del cáncer puede englobar incrementar la función de U1 usando, por ejemplo, agonistas de U1 de la invención para elevar los niveles o actividad de U1 con el fin de atenuar el acortamiento del ARNm. Esto se debe a que la invención se basa en el descubrimiento de que la disminución de los niveles de U1 (p. ej., con un oligonucleótido morfolino antisentido frente a U1) aumenta las características celulares oncogénicas, incluyendo la migración y la capacidad invasiva celulares *in vitro*. Por lo tanto, la invención proporciona composiciones y métodos para sobreexpresar U1 con el fin de atenuar la oncogenicidad.

- 20 Se apreciará que los compuestos (p. ej., agonista de U1 o antagonista de U1) de la invención pueden administrarse a un sujeto bien solos, o conjuntamente con otro agente terapéutico. En una realización, para el tratamiento del cáncer, los compuestos de la presente invención pueden usarse en combinación con agentes terapéuticos existentes usados para tratar el cáncer. En algunos casos, los compuestos de la invención pueden usarse en combinación con estos agentes terapéuticos para aumentar el efecto antitumoral del agente terapéutico.

- 25 Con el fin de evaluar la eficacia terapéutica potencial de los compuestos de la invención en combinación con los terapéuticos antitumorales descritos en otro lugar en la presente memoria, estas combinaciones pueden ensayarse para determinar su actividad antitumoral según métodos conocidos en la técnica.

- 30 En un aspecto, la presente invención contempla que los compuestos de la invención pueden usarse en combinación con un agente terapéutico tal como un agente antitumoral incluyendo, pero no limitado a, un agente quimioterapéutico, un agente antiproliferación celular o cualquier combinación de los mismos.

- 35 La invención no debe limitarse a ningún agente quimioterapéutico particular. En lugar de esto, cualquier agente quimioterapéutico puede unirse a los anticuerpos de la invención. Por ejemplo, cualesquiera agentes quimioterapéuticos convencionales de las siguientes clases ejemplares no limitantes se incluyen en la invención: agentes alquilantes; nitrosoureas; antimetabolitos; antibióticos antitumorales; alcaloides de plantas; taxanos; agentes hormonales; y agentes misceláneos.

- 40 Los agentes alquilantes se denominan así debido a su capacidad de añadir grupos alquilo a muchos grupos electronegativos bajo las condiciones presentes en las células, interfiriendo de esta manera con la replicación del ADN para prevenir que las células cancerosas se reproduzcan. La mayoría de los agentes alquilantes no son específicos del ciclo celular. En aspectos específicos, detienen el crecimiento tumoral mediante la reticulación de bases de guanina en las cadenas de doble hélice del ADN. Los ejemplos no limitativos incluyen busulfán, carboplatino, clorambucilo, cisplatino, ciclofosfamida, dacarbazina, ifosfamida, hidroclicloruro de mecloretamina, melfalán, procarbazona, tiotepa, y mostaza de uracilo.

- 45 Los antimetabolitos previenen la incorporación de bases en el ADN durante la fase de síntesis (S) del ciclo celular, prohibiendo el desarrollo y división normales. Los ejemplos no limitativos de antimetabolitos incluyen fármacos tales como 5-fluorouracilo, 6-mercaptopurina, capecitabina, arabinósido de citosina, floxuridina, fludarabina, gemcitabina, metotrexato, y tioguanina.

Hay una variedad de antibióticos antitumorales que previenen generalmente la división celular interfiriendo con enzimas necesarias para la división celular o alterando las membranas que rodean las células. En esta clase se

incluyen las antraciclinas, tales como doxorubicina, que actúan para prevenir la división celular disruptiendo la estructura del ADN y terminando su función. Estos agentes no son específicos del ciclo celular. Los ejemplos no limitativos de antibióticos antitumorales incluyen dactinomicina, daunorrubicina, doxorubicina, idarrubicina, mitomicina-C, y mitoxantrona.

- 5 Los alcaloides de plantas inhiben o paran la mitosis o inhiben las enzimas que previenen que las células hagan proteínas necesarias para el crecimiento celular. Los alcaloides de plantas usados frecuentemente incluyen vinblastina, vincristina, vindesina, y vinorelbina. Sin embargo, no debe considerarse que la invención esté limitada solamente a estos alcaloides de plantas.

- 10 Los taxanos afectan las estructuras celulares denominadas microtúbulos que son importantes para las funciones celulares. En el crecimiento celular normal, los microtúbulos se forman cuando una célula empieza a dividirse, pero una vez la célula para de dividirse, los microtúbulos se desensamblan o destruyen. Los taxanos evitan que los microtúbulos se degraden de manera que las células cancerosas se agregan con los microtúbulos hasta el punto de que no pueden crecer ni dividirse. Los taxanos ejemplares no limitantes incluyen paclitaxel y docetaxel.

- 15 Los agentes hormonales y fármacos semejantes a hormonas se utilizan para determinados tipos de cáncer, incluyendo, por ejemplo, leucemia, linfoma, y mieloma múltiple. Se emplean frecuentemente con otros tipos de fármacos de quimioterapia para aumentar su efectividad. Las hormonas sexuales se usan para alterar la acción o producción de hormonas femeninas o masculinas y se usan para ralentizar el crecimiento de los cánceres de mama, próstata, y endometrial. La inhibición de la producción (inhibidores de aromataza) o acción (tamoxifeno) de estas hormonas puede usarse frecuentemente como un adyuvante de la terapia. Algunos otros tumores también son dependientes de hormonas. El tamoxifeno es un ejemplo no limitativo de agente hormonal que interfiere con la actividad del estrógeno, que promueve el crecimiento de las células de cáncer de mama.

Los agentes misceláneos incluyen quimioterapéuticos tales como bleomicina, hidroxurea, L-asparaginasa, y procarbazona que también son útiles en la invención.

- 25 Un agente antiproliferación celular puede definirse además como un agente inductor de la apoptosis o un agente citotóxico. El agente inductor de la apoptosis puede ser una granzima, un miembro de la familia Bcl-2, citocromo C, una caspasa, o una combinación de los mismos. Las granzimas ejemplares incluyen granzima A, granzima B, granzima C, granzima D, granzima E, granzima F, granzima G, granzima H, granzima I, granzima J, granzima K, granzima L, granzima M, granzima N, o una combinación de las mismas. En otros aspectos específicos, el miembro de la familia Bcl-2 es, por ejemplo, Bax, Bak, Bcl-Xs, Bad, Bid, Bik, Hrk, Bok, o una combinación de los mismos.

- 30 En aspectos adicionales, la caspasa es caspasa-1, caspasa-2, caspasa-3, caspasa-4, caspasa-5, caspasa-6, caspasa-7, caspasa-8, caspasa-9, caspasa-10, caspasa-11, caspasa-12, caspasa-13, caspasa-14, o una combinación de las mismas. En aspectos específicos, el agente citotóxico es TNF- α , gelonina, Prodigiosina, una proteína inhibidora de ribosomas (RIP), exotoxina de Pseudomonas, Toxina B de Clostridium difficile, VacA de Helicobacter pylori, YopT de Yersinia enterocolitica, Violaceína, ácido dietilentriaminopentaacético, irofulveno, Toxina de la Difteria, mitogilina, ricina, toxina botulínica, toxina del cólera, saporina 6, o una combinación de los mismos.

- 40 En algunas realizaciones, una cantidad efectiva de un compuesto de la invención y un agente terapéutico es una cantidad sinérgica. Tal y como se usa en la presente memoria, una "combinación sinérgica" o una "cantidad sinérgica" de un compuesto de la invención y un agente terapéutico es una combinación o cantidad que es más efectiva en el tratamiento terapéutico o profiláctico de una enfermedad que la mejora incremental en el resultado del tratamiento que podría predecirse o esperarse de una combinación meramente aditiva de (i) el beneficio terapéutico o profiláctico del compuesto de la invención cuando se administra a la misma dosificación como una monoterapia y (ii) el beneficio terapéutico o profiláctico del agente terapéutico cuando se administra a la misma dosificación como una monoterapia.

Composiciones farmacéuticas y terapias

- 45 Las composiciones de la invención pueden administrarse a un sujeto en una composición farmacéutica. Tal y como se usa en la presente memoria, el término "vehículo farmacéuticamente aceptable" se pretende que incluya cualquiera y todos los disolventes, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y retardantes de la absorción, y semejantes, compatibles con la administración farmacéutica. El uso de dichos medios y agentes para sustancias farmacéuticamente activas es muy conocido en la técnica. Excepto en el caso en el que cualquier medio convencional o agente sea incompatible con el compuesto activo, se contempla el uso del mismo en las composiciones. También pueden incorporarse en las composiciones compuestos activos suplementarios.

Aunque la descripción de composiciones farmacéuticas proporcionadas en la presente memoria está dirigida principalmente a composiciones farmacéuticas que son adecuadas para la administración ética a seres humanos, el experto en la técnica entenderá que dichas composiciones son adecuadas generalmente para la administración a

sujetos de todos los tipos. Se entiende bien la modificación de las composiciones farmacéuticas adecuadas para la administración a seres humanos con el fin de hacer que las composiciones sean adecuadas para la administración a varios animales, y el farmacólogo veterinario experto puede diseñar y realizar dicha modificación meramente con experimentación ordinario, si es necesario. Los sujetos para los que se contempla la administración de las composiciones farmacéuticas de la invención incluyen, pero no están limitados a, seres humanos y otros primates, mamíferos incluyendo mamíferos comercialmente relevantes tales como primates no humanos, ganado, cerdos, caballos, ovejas, gatos y perros.

Las composiciones farmacéuticas que son útiles en los métodos de la invención pueden prepararse, envasarse o venderse en formulaciones adecuadas para la administración oftálmica, oral, parenteral, intranasal, bucal u otra ruta de administración. Otras formulaciones contempladas incluyen nanopartículas proyectadas, preparaciones liposomales, eritrocitos resellados que contienen el ingrediente activo, y formulaciones basadas en inmunología.

Una composición farmacéutica de la invención puede prepararse, envasarse, o venderse a granel, como una única dosis unitaria o como una pluralidad de únicas dosis unitarias. Tal y como se usa en la presente memoria, una "dosis unitaria" es una cantidad discreta de la composición farmacéutica que comprende una cantidad predeterminada del ingrediente activo. La cantidad del ingrediente activo es generalmente igual a la dosificación del ingrediente activo que se administraría a un sujeto o una fracción conveniente de dicha dosificación tal como, por ejemplo, la mitad o un tercio de dicha dosificación.

Las cantidades relativas del ingrediente activo, del vehículo farmacéuticamente aceptable y cualesquiera ingredientes adicionales en una composición farmacéutica de la invención variarán, dependiendo de la identidad, tamaño y afección del sujeto tratado y dependiendo además de la ruta por la que se va a administrar la composición. Como ejemplo, la composición puede comprender entre el 0,1 % y el 100 % (p/p) del ingrediente activo.

Además del ingrediente activo, una composición farmacéutica de la invención puede comprender además uno o más agentes farmacéuticamente activos adicionales.

Las formulaciones de liberación controlada o sostenida de una composición farmacéutica de la invención pueden prepararse usando tecnología convencional.

Tal y como se usa en la presente memoria, la "administración parenteral" de una composición farmacéutica incluye cualquier ruta de administración caracterizada por la rotura física de un tejido de un sujeto y la administración de la composición farmacéutica a través de la rotura en el tejido. La administración parenteral incluye así, pero no está limitada a, la administración de una composición farmacéutica por inyección de la composición, por aplicación de la composición a través de una incisión quirúrgica, por aplicación de la composición a través de una herida no quirúrgica que penetra en el tejido y semejantes. En particular, se contempla que la administración parenteral incluye, pero no está limitada a, inyección intraocular, intravítrea, subcutánea, intraperitoneal, intramuscular, intraesternal, intratumoral y técnicas de infusión dialíticas renales.

Las formulaciones de una composición farmacéutica adecuadas para administración parenteral comprenden el ingrediente activo combinado con un vehículo farmacéuticamente aceptable, tal como agua estéril o disolución salina isotónica estéril. Dichas formulaciones pueden prepararse, envasarse o venderse en una forma adecuada para la administración en bolo o para la administración continua. Las formulaciones inyectables pueden prepararse, envasarse, o venderse en forma de dosificación unitaria, tal como en ampollas o en contenedores de múltiples dosis que contienen un conservante. Las formulaciones para la administración parenteral incluyen, pero no están limitadas a, suspensiones, disoluciones, emulsiones en vehículos grasos o acuosos, pastas y formulaciones de liberación sostenida implantables o biodegradables. Dichas formulaciones pueden comprender además uno o más ingredientes adicionales incluyendo, pero no limitado a, agentes de suspensión, estabilizantes o dispersantes. En una realización de una formulación para administración parenteral, el ingrediente activo se proporciona en forma seca (es decir, en polvo o granular) para su reconstitución con un vehículo adecuado (p. ej., agua estéril sin pirógenos) antes de la administración parenteral de la composición reconstituida.

Las composiciones farmacéuticas pueden prepararse, envasarse, o venderse en la forma de una suspensión o disolución inyectable estéril acuosa o aceitosa. Esta suspensión o disolución puede formularse según la técnica conocida, y puede comprender, además del ingrediente activo, ingredientes adicionales tales como los agentes dispersantes, agentes humectantes o agentes de suspensión descritos en la presente memoria. Dichas formulaciones inyectables estériles pueden prepararse usando un diluyente o disolvente no tóxico parenteralmente aceptable, tal como agua o 1,3-butano diol, por ejemplo. Otros diluyentes y disolventes aceptables incluyen, pero no están limitados a, disolución de Ringer, disolución de cloruro de sodio isotónica y aceites fijados tales como mono o diglicéridos sintéticos. Otras formulaciones administrables parenteralmente que son útiles incluyen las que comprenden el ingrediente activo en forma microcristalina, en una preparación liposomal o como un componente de un sistema polimérico biodegradable. Las composiciones para liberación sostenida o implante pueden comprender materiales

poliméricos o hidrófobos farmacéuticamente aceptables tales como una emulsión, una resina de intercambio iónico, un polímero poco soluble o una sal poco soluble.

Las formulaciones adecuadas para administración nasal pueden comprender, por ejemplo, de aproximadamente tan poco como el 0,1 % (p/p) y tanto como el 100 % (p/p) del ingrediente activo, y pueden comprender además uno o más de los ingredientes adicionales descritos en la presente memoria.

Una composición farmacéutica de la invención puede prepararse, envasarse o venderse en una formulación adecuada para la administración bucal. Dichas formulaciones pueden estar, por ejemplo, en la forma de comprimidos o pastillas para chupar preparadas usando métodos convencionales, y pueden, por ejemplo, el 0,1 al 20 % (p/p) de ingrediente activo, comprendiendo el equilibrio una composición que se puede disolver oralmente o degradable y, opcionalmente, uno o más de los ingredientes adicionales descritos en la presente memoria. Alternativamente, las formulaciones adecuadas para administración bucal pueden comprender un polvo o una disolución o suspensión aerosolizada o atomizada que comprende el ingrediente activo. Dichas formulaciones en polvo, aerosolizadas, o aerosolizadas, cuando se dispersan, tienen preferiblemente un tamaño de partícula o de gotita promedio en el rango de aproximadamente 0,1 a aproximadamente 200 nanómetros, y pueden comprender además uno o más de los ingredientes adicionales descritos en la presente memoria.

Tal y como se usa en la presente memoria, los "ingredientes adicionales" incluyen, pero no están limitados a, uno o más de los siguientes: excipientes; agentes tensioactivos; agentes dispersantes; diluyentes inertes; agentes de granulación y disgregación; agentes de unión; agentes lubricantes; agentes edulcorantes; agentes saboríferos; agentes colorantes; conservantes; composiciones fisiológicamente degradables tales como gelatina; vehículos y disolventes acuosos; vehículos y disolventes aceitosos; agentes de suspensión; agentes dispersantes o humectantes; demulcentes; tampones; sales; agentes espesantes; rellenos; agentes emulsionantes; antioxidantes; antibióticos; agentes antifúngicos; agentes estabilizantes; y materiales poliméricos o hidrófobos farmacéuticamente aceptables. Otros "ingredientes adicionales" que pueden estar incluidos en las composiciones farmacéuticas de la invención son conocidos en la técnica y se describen, por ejemplo, en Remington's Pharmaceutical Sciences (1985, Genaro, ed., Mack Publishing Co., Easton, Pa.).

Los métodos terapéuticos y profilácticos de la invención engloban así el uso de composiciones farmacéuticas que comprenden un oligonucleótido de la invención que altera el corte y empalme para llevar a la práctica los métodos de la invención. La dosificación precisa administrada variará dependiendo de cualquier número de factores, incluyendo, pero no limitado a, el tipo de animal y el tipo de estado de enfermedad que se está tratando, la edad del animal y la ruta de administración.

El compuesto puede administrarse a un sujeto tan frecuentemente como varias veces al día, o puede administrarse menos frecuentemente, tal como una vez al día, una vez a la semana, una vez cada dos semanas, una vez al mes o incluso menos frecuentemente, tal como una vez cada varios meses o incluso una vez al año o menos. La frecuencia de la dosis será fácilmente evidente para el experto en la técnica y dependerá de cualquier número de factores, tales como, pero no limitado a, el tipo y gravedad de la enfermedad que se está tratando, el tipo y edad del sujeto, etc. Las formulaciones de las composiciones farmacéuticas descritas en la presente memoria pueden prepararse por cualquier método conocido o desarrollado a partir de ahora en la técnica de la farmacología. En general, dichos métodos preparatorios incluyen la etapa de asociar el ingrediente activo con un vehículo o uno o más de otros ingredientes accesorios, y después, si es necesario o deseable, dar forma o envasar el producto en una unidad deseada de una o múltiples dosis.

Ejemplos experimentales

La invención se describe adicionalmente con detalle por referencia a los siguientes ejemplos experimentales. Estos ejemplos se proporcionan solo para propósitos de ilustración, y no se pretende que sean limitantes a no ser que se especifique otra cosa. Así, no debe considerarse de ninguna forma que la invención está limitada a los siguientes ejemplos, sino que debe considerarse que engloba cualquiera y todas las variaciones que resultarán evidentes como resultado de la enseñanza proporcionada en la presente memoria.

Ejemplo 1: la RNPsn U1 determina la longitud del ARNm y regula la expresión de isoformas

La RNPsn U1 (U1), además de su papel en el corte y empalme, protege a los pre-ARNm frente a la terminación prematura drástica por escisión y poliadenilación (PCPA) en señales de poliadenilación crípticas (PAS) en intrones. Aquí, para definir los parámetros implicados en PCPA y su supresión, se concibió una estrategia (HIDE-seq) para seleccionar y secuenciar solo los transcritos expresados diferencialmente, identificar los cambios que ocurren después de la disminución de U1 a varios niveles y en diferentes organismos. La información de la secuencia obtenida de HIDE-seq proporcionó mapas de PCPA en todo el genoma y estos, junto con experimentos directos, revelaron que la supresión por U1 de PCPA no es solo esencial para proteger a los transcritos nacientes, sino que también es un mecanismo de regulación de la expresión génica global. Inesperadamente, la posición de PCPA varió ampliamente

con el grado de disminución de U1, tendiendo a usar los PAS más próximos con mayor reducción. Esto rindió ARNm con regiones no traducidas en 3' (3' UTR) más cortas e isoformas con corte y empalme alternativo que resultan del uso de sitios de poliadenilación alternativos (APA) más próximos, característico de las células inmunes activadas, neuronales, y cancerosas (Flavell et al., 2008, *Neuron* 60:1022-1038; Mayr y Bartel, 2009, *Cell* 138:673-684; Niibori et al., 2007, *Neurosci Res* 57:399-410; Sandberg et al., 2008, *Science* 320:1643-1647). Se demuestra que una disminución de U1 puede recapitular dichos cambios específicos en el ARNm que ocurren durante la activación neuronal. De hecho, se muestra que la rápida regulación al alza transcripcional durante la activación neuronal es una condición fisiológica que crea escasez de U1 respecto a los transcritos nacientes. Además, la sobreexpresión de U1 inhibe los acortamientos de firma de ARNm de neuronas activadas. Se sugiere que mediante la determinación del grado de supresión de PCPA, los niveles de U1 juegan un papel clave en el uso de PAS y, por lo tanto, en la longitud de los ARNm. Se propone un modelo mediante el cual U1 se une a pre-ARNm nacientes cotranscripcionalmente para explicar cómo la escasez de U1 da lugar a una pérdida correspondiente de la supresión de PAS distal de la maquinaria de escisión y poliadenilación que está asociada con el complejo de elongación de la transcripción (TEC) de la ARN polimerasa II (polII) (Das et al., 2007, *Mol Cell* 26:867-881; Hirose y Manley, 1998, *Nature* 395:93-96; McCracken et al., 1997, *Nature* 385:357-361). Los experimentos adicionales son consistentes con la explicación de que PCPA cotranscripcional contrarrestada por la asociación de U1 con transcritos nacientes, un proceso denominado en la presente memoria como *telescripting*, asegura la integridad del transcriptoma y regula la longitud de los ARNm.

Se describen ahora los materiales y métodos empleados en estos experimentos.

Oligonucleótidos morfolino antisentido

Loa AMO de U1 son iguales en ratones y seres humanos. Esta secuencia y la del AMO de U1 de *Drosophila* (dU1) se listan en otro lugar en la presente memoria.

Tratamientos celulares

Las células HeLa, NIH/3T3, y S2 se transfectaron por electroporación usando un Genepulser (Bio-Rad) o Nucleofector (Amaxa) con control o morfolino antisentido (AMO) frente a U1 a 0,25, 1,0 y 15 nmoles durante 8 hrs como se describe (Kaida et al., 2010, *Nature* 468:664-668). La transfección de HeLa con AMO específicos de gen (15 nmoles/7,5 µM) frente a PAS, sitio de corte y empalme en 5' y 3' se realizó de la misma manera. Todas las secuencias de los AMO sintetizados por Gene Tools se listan en otro lugar en la presente memoria. El tratamiento con SSA fue 100 ng/ml durante 8 hrs. Las células PC12 y MN-1 se estimularon con forskolina 20 µM y/o KCl 50 mM durante 3 hrs (Impey et al., 1998, *Neuron* 21:869-883) o se transfectaron con AMO de U1 durante 8 hrs con Nucleofector de Amaxa.

Preparación de la biblioteca HIDE-Seq

Se prepararon bibliotecas de ADNc sustractivas usando el kit de sustracción de ADNc PCR-Select de Clontech (Diatchenko et al., 1996, *Proc Natl Acad Sci USA* 93:6025-6030; Gurskaya et al., 1996, *Anal Biochem* 240:90-97; Lukyanov et al., 1995, *Anal Biochem* 229:198-202) con varias modificaciones listadas más adelante. El ARN poli(A) se sometió a transcripción inversa con hexámeros aleatorios y cebadores personalizados oligo(dT) que contenían sitios de restricción con extremos romos para preparar conjuntos de ADNc individuales. Después de la síntesis de la segunda cadena, los ADNc ds se digirieron cada uno separadamente con *RsaI*, *HaeIII*, o *AluI*, para producir bibliotecas independientes de las mismas muestras. Cada muestra (AMO control o AMO de U1) sirvió como el ensayador o la referencia (es decir, sustracciones directa e inversa). Las etapas de ligación de adaptadores, hibridación, y PCR primaria (Figura 1A, gel de la izquierda) se hicieron como se describe en otro lugar en la presente memoria. Se realizó PCR anidada (Figura 1A, gel de la derecha) en cada biblioteca con cebadores que contenían adaptadores 454, códigos de barras, y moléculas conectoras. Las bibliotecas directas e inversas se multiplexaron, se sometieron a PCR de emulsión, y se secuenciaron bidireccionalmente usando química de Titanio 454.

Matrices en mosaico genómicas, RT-PCR, en tiempo real y 3' RACE

Las matrices se realizaron en triplicado para los cromosomas humanos 5, 7 y 16 (Kaida et al., 2010, *Nature* 468:664-668). Todos los cebadores, métodos de preparación de ARN, condiciones, y tipos celulares se listan en otro lugar en la presente memoria. La isoforma de ARNm y los niveles de ARNsn de U1 determinados por qPCR se normalizaron respecto a G6PDH y ARNr 5s, respectivamente.

Mini-Gen, transfecciones de U1 mutante y sobreexpresión de U1

El minigen NR3C1 y las mutaciones en el sitio de corte y empalme en 5' y PAS contenidas en él se han descrito (Kaida et al., 2010, *Nature* 468:664-668). Para duplicar el PAS, el fragmento *EcoRV-StuI* del plásmido NR3C1 se reinsertó en el minigen original en el sitio *EcoRV*. La modificación en el dominio de unión del sitio de corte y empalme en 5' de U1 y las concentraciones de plásmidos se describen en otro lugar en la presente memoria. Los productos de 3'RACE se digirieron con *HindIII* para distinguir las bandas de PCPA y ARNm. Las PC12 se transfectaron con ARNsn de U1

en un vector de expresión pSiren-RetroQ (Clontech) dirigido por el promotor nativo de U1. Se usó RT-qPCR usando sondas específicas para una diferencia de 1 pb en U1 transfectada para determinar el grado de sobreexpresión.

Marcaje del ARNm con uridina tritiada

5 Las células PC12 se pulsaron con 50 μ Ci de 5'6'-³H uridina en 500 μ l de medio con 0,04 μ g/ml de actinomicina D durante 30 min antes de recogerlas. El ARN total se aisló por Trizol y el ARN poli(A) se seleccionó en lechos Oligotex (Qiagen). La radiactividad en cada fracción se determinó por conteo de centelleo y las relaciones de ARNm a ARN total se normalizaron respecto a la concentración de ARN total determinada en un espectrofotómetro Nanodrop.

Secuenciación de alto rendimiento

10 La secuenciación 454 con titanio se realizó en la Instalación de Secuenciación de la Universidad de Pensilvania o en MacroGen (Seúl, Corea) según las recomendaciones del fabricante.

Análisis de los datos y construcción de una estructura (*pipeline*) bioinformática

15 Las lecturas de secuencia se filtraron para cebadores correctos, se categorizaron por códigos de barras y se alinearon con el genoma humano (hg18) usando GMAP (Wu y Watanabe, 2005 Bioinformatics 21: 1859-75) con parámetros por defecto. El mejor alineamiento para una lectura se basó en la cobertura (es decir, el porcentaje de la lectura alineada con el genoma), y en el evento de un empate en la puntuación de cobertura, se eligió el alineamiento con la mayor identidad de secuencia. Las lecturas con múltiples alineamientos con el genoma, que tienen la misma cobertura e identidad se desecharon, igual que las lecturas con menos del 90 % de cobertura/90 % de identidad. Se obtuvieron resultados de alineamiento comparables con BLAT [www.genome.org/cgi/content/abstract/12/4/656]. Los alineamientos se aglomeraron entonces y las regiones del genoma afectadas se identificaron para cada sustracción, es decir, bien Control-U1 o U1-Control. Las regiones en ambas regiones de sustracción que se alineaban con la misma localización genómica se consideraron fondo y se desecharon. Las regiones remanentes afectadas de forma única en cada sustracción se anotaron entonces usando las anotaciones de transcrito canónicas de Refseq hg18. En el raro evento de múltiples isoformas de transcrito canónico para un único gen, se eligió el transcrito más largo como el transcrito representativo para un gen. Se generan archivos Bed para la visualización de los datos en el navegador de genoma de la UCSC. El uso de la estructura de análisis de los datos está disponible en (www.upenn.edu/dreyfusslab).

Identificación de lecturas poli(A), gráficos de densidad y cálculos

30 Las lecturas de HIDE-seq en las sustracciones U1-Control con colas de poli(A) no encontradas en la secuencia genómica se seleccionaron como sigue: para todas las lecturas, se retiraron el código de barras (no. 2) y cebadores, los datos se recortaron para calidad (puntuación <20) y las secuencias se alinearon con los genomas de referencia respectivos. Las secuencias que terminaban en cadenas de 6 o más A que no se alineaban con el genoma, o para las que la secuencia de alineamiento en 5' no estaba en una 3'UTR anotada, se categorizaron y se confirmaron por inspección visual. Las coordenadas genómicas y las secuencias de PCPA definitivas encontradas en seres humanos (n=71), ratón (n=70) y Drosophila (n=60) se encuentran en la Figura 9, Figura 10, y Figura 11. Las puntuaciones de PAS se generaron usando la herramienta de predicción polya_svm en: polya.umd.edu/polyadb/. Las longitudes y distancias génicas respecto al TSS, el sitio de corte y empalme en 5' y sitio de corte y empalme en 3' se representaron en R. Estos y las cantidades de secuencia intrónica se calcularon usando la base de datos de transcritos canónicos (Rhead et al., 2010 Nucleic Acids Res. 38: D613-D619).

Clasificación de cambios en el transcriptoma

40 Para evaluar la prevalencia relativa de los patrones observados (Figura 2B), se aplicó una estrategia bioinformática no sesgada para clasificar los cambios en el transcriptoma. Cada transcrito de pre-ARNm se trata como una secuencia de bloques de exones e intrones alternante y el alineamiento de una o más lecturas en cualquier sustracción indica la presencia de un cambio único. Los transcritos se pusieron inicialmente en cuatro clases usando una estrategia heurística, basada en que el patrón de bloques consecutivos exón e intrón estaba "arriba" o "abajo". La clase 1 representa transcritos que muestran al menos un exón o intrón que está arriba, mientras ninguno está abajo. La clase 2 tiene los criterios opuestos. Los transcritos de la clase 3 muestran un patrón Z, que consiste en una serie de uno o más exones o intrones que están "arriba", seguido de uno o más exones que está abajo. La clase 4 contiene todos los "demás" patrones que no se encuentran en una categoría definida.

50 Se consiguió una subclasificación adicional de estos con la siguiente estrategia de optimización, comparando el transcrito observado con patrones ideales para cada uno. La clase 1 se subdividió en: todos los exones solo "arriba", todos los exones "arriba" con retención de intrones, exones y/o intrones en 5' "arriba" sin cambio en el extremo 3', que se refiere como un patrón 7, retención de intrones único o múltiple sin sesgo en 5', u "Otro". Conceptualmente, el caso de todos los exones "arriba" puede verse como una línea recta con amplitud normalizada 1. Un 7 puede verse como una función escalonada que tiene amplitud de 1 en los números de exones 1 a k, donde k es idealmente igual a la

mitad del número de exones, y amplitud 0 en los números de exones k+1 a n, donde n es el número de exones en el transcrito. Por lo tanto, la clasificación se redujo para ajustarse a una función sigmoidea o escalonada de los datos y estimar k mediante regresión de mínimos cuadrados. Si k era menor de 0,75 n, entonces el transcrito se clasifica como un 7, de otra forma el transcrito se clasificó como todos los exones “arriba” sin retención de intrones. Si había demasiados pocos cambios en los exones para tomar una decisión, entonces el transcrito se clasificó como “Otro”. La subdivisión de la clase 2 en “Todos los Exones Abajo”, “L” u “Otro” se realizó de una forma similar. Las clasificaciones se verificaron por inspección visual de los archivos bed en el navegador de genoma USCS.

Preparación de ARN, RT-PCR y 3' RACE

El ARN total se preparó con el kit RNeasy (Qiagen) y el ARN poli(A) se aisló del ARN total con el kit de ARNm Oligotex (Qiagen). Se realizó RT-PCR estándar con ADNc preparado por Advantage RT-for-PCR (Clontech) a partir de ARN total de células transfectadas. Se realizó RT-PCR cuantitativa en tiempo real en un Applied Biosystems Fast 7500 usando química de tinción SYBR verde. Las veces de cambio se normalizaron respecto a los niveles de ARNm de actina. Para la cuantificación de los ARNsn, se aisló el ARN total de líneas celulares humanas (HeLa, 293T, U20S, y fibroblastos primarios), de ratón (fibroblastos primarios, músculo y astrocitos; MN1 y C2C12) y de Drosophila (S2 y Kc167) usando el kit de aislamiento de miARN mirVana (Ambion) según recomienda el fabricante. Se usaron 10 ng (o 2 ng para U1) de ARN total para las reacciones de amplificación de RT lineal usando el kit Advantage RT-for-PCR (Clontech), con cebadores inversos específicos para cada ARNsn. Se usó el 4 % del producto de ADNc como entrada en las mediciones de qPCR en tiempo real usando un sistema de PCR en tiempo real 7900HT Fast (ABI). Se generaron curvas estándar con diluciones seriadas de cantidades conocidas de producto de PCR ARNsn purificado y se produjeron para cada experimento para asegurar la exactitud. Para 3' RACE, el ARN se transcribió de forma inversa usando un cebador [oligo(dT)-XbaIKpnIHindIII], seguido de PCR anidada usando cebadores directos específicos de gen o específicos de plásmido distintos y un cebador inverso común, XbaIKpnIHindIII, durante 20 ciclos cada uno.

Mutación de genes de ARNsn de U1.

La secuencia de reconocimiento del sitio de corte y empalme en 5' de tipo salvaje de U1 (5'-ACTTACCTG-3'; SEQ ID NO: 1) se mutó a las siguientes secuencias para la Figura 5C: U1-A (5'-CTGCTGTTG-3'; SEQ ID NO: 2), U1-B (5'-TGAATGGAC-3'; SEQ ID NO: 3), U1-C (5'-CTACAGATG-3'; SEQ ID NO: 4), y U1-D (5'-CACCCCTAC-3'; SEQ ID NO: 5).

Secuencias de oligonucleótidos morfolinos antisentido y cebadores de PCR

Morfolinos Antisentido:	Secuencia (SEQ ID NO)
AMO Control	CCTCTTACCTCAGTTACAATTTATA (SEQ ID NO: 6)
AMO de U1	GGTATCTCCCCTGCCAGGTAAGTAT (SEQ ID NO: 7)
AMO de U2	TGATAAGAACAGATACTACACTTGA (SEQ ID NO: 8)
AMO de U6atac	AACCTTCTCTCCTTTTCATACAACAC (SEQ ID NO: 9)
AMO de U12	TCGTTATTTTCCTTACTCATAAGT (SEQ ID NO: 10)
AMO de dU1	GTAAACCTCTACGCCAGGTAAGTAT (SEQ ID NO: 11)
AMO de dU2	GATAAGAACAGATACTACACTTTGA (SEQ ID NO: 12)
AMO de dU6atac	AACTTGCTCTCCTTCCAAACAACAC (SEQ ID NO: 13)
AMO de dU12	GATTGGTTTTCCTTACTCATTAGTT (SEQ ID NO: 14)
AMO de 3'ss de UBAP2L	GGATATTTTCTTCTTAAAAGGAATC (SEQ ID NO: 15)
AMO de PAS de UBAP2L	AATACCACAGTTTATTGAAGACTAC (SEQ ID NO: 16)

Cebadores de HIDE-Seq:

Los cebadores híbridos consisten en secuencias 454 A o B (itálica), códigos de barras (subrayado) y secuencias adaptadoras de PCR de supresión (negrita).

Biblioteca Control-U1:

DIR 5'-CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT**GCTCGAGCGGCCGCCCGGGCAGGT**-3' (SEQ ID NO: 17)

INV 5'-CTATGCGCCTTGCCAGCCCGCTCAGACGAGTGCGT**GGCAGCGTGGTCCGCGGCCGAGGT**-3' (SEQ ID NO: 18)

Biblioteca U1-Control:

DIR 5'-CGTATCGCCTCCCTCGCGCCATCAGACGCTCGAC**GCTCGAGCGGCCGCCCGGGCAGGT**-3' (SEQ ID NO: 19)

INV 5'-CTATGCGCCTTGCCAGCCCGCTCAGACGCTCGACA**GGCAGCGTGGTCGCGGCCGAGGT**-3' (SEQ ID NO: 20)

Cebadores de 3' RACE:

Los cebadores NR3C1 se listan en Kaida et al., 2010.

5 Para detectar PCPA en BASP1:

BASP1 int1-1: CAGTCTGCCATTGTGACGTG (SEQ ID NO: 21)

BASP1 int1-2: CCAATGGGAGAGCCCTAAAC (SEQ ID NO: 22)

Para detectar poliadenilación en la 3' UTR:

BASP1 3UTR-1: CTGCAATGGGAGTTGGGAG (SEQ ID NO: 23)

10 BASP1 3UTR-2: GATGCGTTTGATTCTGCCAC (SEQ ID NO: 24)

ChicEx3 DIR1: GAGGAGCTCTCCAACTGATCAGCG (SEQ ID NO: 25)

ChicEx3 DIR2: CTGATCAGCGGCTTTGACCA (SEQ ID NO: 26)

Ten-mEx1 DIR1: CGAGTATGAGTCGACTCTAGACTGCCG (SEQ ID NO: 27)

Ten-mEx1 DIR2: CTCTAGACTGCCGCGACGTGG (SEQ ID NO: 28)

15 Cebadores inversos universales:

Oligo dT18-XbaKpnBam: CTGATCTAGAGGTACCGGATCCTTTTTTTTTTTTTTTTTT (SEQ ID NO: 29)

XbaKpnBam: CTGATCTAGAGGTACCGGATCC (SEQ ID NO: 30)

Cebadores de RT-PCR:

Para detectar ARNm sometido a corte y empalme:

20 BASP1Ex1D: GCTAACTCAGGGGCTGCATAGGCAC (SEQ ID NO: 31)

BASP1EX2I: CGTCCTTCTCGCCCTCCTTCTCCTC (SEQ ID NO: 32)

Para detectar región en 5' de PAS en intrón (IR1):

BASP1USID: GGGTGGCTTTCTTAATTTGCATC (SEQ ID NO: 33)

BASP1USII: GATGATTTCCATTGCTGCCACG (SEQ ID NO: 34)

25 Para detectar región en 3' de PAS en intrón (IR2):

BASP1DSID: GTCTTGAAACACAGCAGTGCCAG (SEQ ID NO: 35)

BASP1DSII: CTCAGCACTGACGGCCTTGTTGAGT (SEQ ID NO: 36)

Homer1 DIRcorto/largo: CAGACGATGAGAGAACACCCGATG (SEQ ID NO: 37)

Homer1 INVcorto: TGGCACCTCTGTGGGCCTGTGGC (SEQ ID NO: 38)

30 Homer1 INVlargo: CTGTTGCTTCCACTGCTTCACATTGGC (SEQ ID NO: 39)

Dab1 DIRcorto/largo: AAGGATAAGCAGTGTGAACAAG (SEQ ID NO: 40)

Dab1 INVcorto: TCTAGATCTCCCATCACGGC (SEQ ID NO: 41)

Dab1 INVlargo: CAGCAGTGCCGAAAGACATA (SEQ ID NO: 42)

G6DPH DIR: CGCCTATCAGTCTGTCCCTGGACTCC (SEQ ID NO: 43)

35 G6DPH INV: CCTTCCTTTGGTAAGGAGGCTGCCC (SEQ ID NO: 44)

CG5168 Ex1 DIR: CCGGAATTGCTGAGTAAGCTGG (SEQ ID NO: 45)

CG5168 Ex2 INV: GGAAGGCATATACTGGCAAATGC (SEQ ID NO: 46)

Gint3 Ex3 DIR: CCTTTAGCTCCTCGAATTTAACAG (SEQ ID NO: 47)

Gint3 Ex4 INV: CCTTGATGAACTTCCTCAGAGGAC (SEQ ID NO: 48)

RAN Ex3 DIR: GGTGGGTGATGGTGGTACTGG (SEQ ID NO: 49)

5 RAN Ex5 INV: GATTTGCCTTCACTTTCTGTC (SEQ ID NO: 50)

POLR1C Ex3 DIR: GTGTGGATGTAGTACACATGGATGAAAAC (SEQ ID NO: 51)

POLR1C Ex4 INV: TTATTGTACACCAGGACCTTCTCCACAG (SEQ ID NO: 52)

POLR1C Ex4 DIR: CAACTATGGCTGTGGAGAAGGTCCTGGT (SEQ ID NO: 53)

POLR1C Ex5 INV: TCTGACCTGGAGACGAACTGTAGAGTATCT (SEQ ID NO: 54)

10 Cebadores de RT-qPCR en tiempo real:

GABPB1 8DIR: CAACAGACATTGCTGAAGAAAC (SEQ ID NO: 55)

GABPB1 8bleedINV: ATTTTGGATGACTGCGGC (SEQ ID NO: 56)

GABPB1 9INV: CTTTAGGAGCTGCTGTGCGATA (SEQ ID NO: 57)

UBAP2L 24DIR1: GCTCCTACCTCTTCCAAGCAG (SEQ ID NO: 58)

15 UBAP2L apa25INV: ACAAACACAGCCCCCAGC (SEQ ID NO: 59)

UBAP2L 24DIR2: TGGTGTGAATGTGAGTGTGAATGC (SEQ ID NO: 60)

UBAP2L 2526INV: GCAAGTTGAAGGAAGCAGCAGG (SEQ ID NO: 61)

ARNsn de U1 humano DIR: TGATCACGAAGGTGGTTTTCC (SEQ ID NO: 62)

ARNsn de U1 humano INV: GCACATCCGGAGTGCAATG (SEQ ID NO: 63)

20 Se describen ahora los resultados de los experimentos.

HIDE-Seq: una estrategia para la secuenciación de alto rendimiento solo de diferencias en el transcriptoma

Para identificar los cambios en el transcriptoma después de la depleción funcional de la RNPsn U1, se prepararon bibliotecas de ADNc a partir de ARN poli(A) de células humanas (HeLa) 8 hr después de la transfección con 15 nmoles de AMO de U1 (deplecionadas de U1) o AMO control (Kaida et al., 2010, Nature 468:664-668). Cada biblioteca de ADNc se digirió separadamente con tres endonucleasas de restricción de 4 pb para producir fragmentos con una longitud uniforme (~250 pb), y se ligaron diferentes oligonucleótidos adaptadores separadamente a los extremos 5' del ADNc experimental para la amplificación posterior. La hibridación sustractiva y la PCR de supresión amplificaron selectivamente solo los transcritos expresados diferencialmente (Diatchenko et al., 1996, Proc Natl Acad Sci USA 93:6025-6030; Gurskaya et al., 1996, Anal Biochem 240:90-97). Se usaron cebadores anidados fusionados a conectores secuenciadores 454 y códigos de barras específicos de muestra para generar amplicones de bibliotecas sustraídas preparados en ambas direcciones U1-Control y Control-U1 para la secuenciación paralela masiva en el mismo pocillo, controlando la variación de muestra a muestra (Figura 1A). Las lecturas de secuencias recíprocas de la sustracción bidireccional potenciaron adicionalmente la definición de los cambios en el transcriptoma.

Se realizó la secuenciación de amplicones de alto rendimiento usando solo 1/4-1/2 de una placa de secuenciación 454 (~300.000 lecturas) para el perfilado extenso (Tabla 1). Cerca del 70 % de las lecturas de HIDE-seq, promediando 150-400 nt, se mapearon de forma no ambigua (90 % de identidad/90% de cobertura) en el genoma (Tabla 1). Las lecturas localizadas en regiones intergénicas y secuencias no únicas para ninguna de las direcciones de sustracción se excluyeron del análisis adicional. Solo una fracción menor (0,03 %) fue de ARN ribosomal comparado con >70 % sin sustracción, lo que confirma la eficiencia del método e indica que la mayoría de las lecturas son informativas (Tabla 1). De forma importante, al obtenerse más lecturas, el número de nuevos genes afectados descubiertos alcanzó una meseta (Figura 1B), como lo hizo la cobertura en un gen, lo que sugiere que se consiguió una cobertura extensa y que permitió una ausencia de lecturas para indicar un cambio de secuencia pequeño o ausente en una localización dada.

Los datos de las matrices en mosaico genómicas (GTA) de alta densidad de los cromosomas humanos 5, 7, y 16 para las células HeLa tratadas con AMO de U1 bajo las mismas condiciones usadas aquí (15 nmoles) proporcionaron un conjunto de datos exhaustivo con el que comparar la metodología HIDE-seq (Kaida et al., 2010, Nature 468:664-668).

Las lecturas de HIDE-seq presentes en un locus genómico dado se interpretaron como que estaban arriba o abajo en las células tratadas con AMO de U1 respecto al control, según si provienen de la sustracción U1-Control o control-U1, respectivamente (Figura 1C). Para GTA, se usó SSA como una referencia para la acumulación de intrones (p. ej., NSUN2 y SKIV2L2) y para facilitar el descubrimiento de PCPA, representado por la acumulación de intrones en 5' que termina abruptamente seguido de una disminución en las señales aguas abajo (p. ej., RFWD3 y CUL1). HIDE-seq identificó fácilmente los mismos cambios en el transcriptoma como GTA (Figura 1C). Respecto a los cromosomas 5, 7 y 16, capturó las acumulaciones de intrones altamente significativas (≥ 2 veces de cambio, ≥ 100 nt, valor $p < 0,01$) detectadas por GTA en el 85 % (189/223) de los genes (Figura 8A), una correspondencia muy alta considerando que los ARN provenían de experimentos biológicos separados. HIDE-seq descubrió 198 genes adicionales con acumulación de intrones en los tres mismos cromosomas, el 82 % de los cuales se confirmó con análisis GTA de baja astringencia, indicando que HIDE-seq es altamente fiable y sensible. Además, HIDE-seq identificó estas y otras diferencias en el genoma, capturando cambios sutiles de secuencia tales como omisión de exones (Figura 8B y Figura 8C). Aunque HIDE-seq no es cuantitativo o necesariamente completo, representa una estrategia alternativa y complementaria a la secuencia del transcriptoma completo, proporcionando una instantánea detallada de las diferencias en el transcriptoma a una fracción del coste. Combinado con una estructura informática simplificada (www.upenn.edu/dreyfusslab), HIDE-seq es una estrategia simple y potente que es ampliamente aplicable con diferentes plataformas de secuenciación (p. ej., Illumina).

Tabla 1. HIDE-seq produjo lecturas de alta calidad de transcritos expresados diferencialmente

	Humanos (15)	Ratón (15)	Drosophila (15)	Humanos (1,0)	Humanos (0,25)
Número total de lecturas con cebadores 454	394.206	255.808	304.454	97.258	74.889
Número total de lecturas alineadas	337.646 (86 %)	243.086 (95 %)	274.098 (90 %)	54.760 (56 %)	39.541 (53 %)
Lecturas alineadas únicamente en el umbral 90/90	232.346 (69 %)	90.056 (37 %)	162.416 (59 %)	50.794 (93 %)	36.733 (93 %)
Número de lecturas 90/90 usadas en anotación	150.361 (65 %)	68.957 (77 %)	113.989 (70 %)	38.566 (76 %)	27.392 (75 %)
Lecturas de fondo, intergénicas, cortas	81.985 (35 %)	21.099 (23 %)	48.427 (30 %)	12.228 (24 %)	9.341 (25 %)
Lecturas de ARN ribosomal	80 (0,03 %)	80 (0,09 %)	39 (0,02 %)	nd	nd

Los resultados de HIDE-seq después del tratamiento con AMO de U1 se resumen para experimentos humanos (Hela), de ratón (3T3) y de Drosophila (S2). Se listan los números totales de lecturas y los porcentajes (entre paréntesis). Las lecturas totales fueron aquellas con cebadores 454 y códigos de barras apropiados. Las lecturas alineadas fueron aquellas que se alinearon con el ensamblaje del genoma de referencia apropiado [Hs: marzo 2006 (NCBI36/hg18); Mm: julio 2007 (NCBI37/mm9); Dm: abril 2006 (BDGP R5/dm3)]. Las lecturas usadas en la anotación se mapearon a una posición en el genoma con una cobertura del 90 % (90 % de la lectura alineada) con 90 % de identidad. Las lecturas no incluidas en la anotación incluyeron lecturas de fondo (idénticas) presentes en ambas direcciones de sustracción, lecturas intergénicas que no mapearon a genes conocidos, y lecturas más cortas de 50 pares de bases. El ARN ribosomal representa las cantidades de ARNr 16S y 28S.

PCPA y su supresión por U1 están conservadas evolutivamente

Se usó HIDE-seq para determinar si PCPA ocurre en organismos divergentes después de la depleción de U1 en células de ratón (3T3) y de Drosophila (S2) (Figura 1B) con AMO específicos para la secuencia del extremo 5' de U1 de cada organismo. Como en los seres humanos, la terminación prematura en intrones fue específica de la RNPsn U1 y no fue una consecuencia de la inhibición del corte y empalme (Figura 8E y Figura 8F). HIDE-seq en HeLa, 3T3, y S2 detectó diferencias de secuencia en 6.548, 5.724, y 3.283 genes, respectivamente, que podría clasificarse en varios patrones (Figura 2B). La acumulación de lecturas de intrones poliadenilados, seguido de una disminución en señales de exones en 3', proporcionó la evidencia más directa de PCPA (Figura 2A, Figura 9, Figura 10, y Figura 11). Este patrón, designado como Z (Figura 2B), demostró que PCPA ocurría típicamente en el intrón 1, con poca o ninguna transcripción más allá de ese punto. Se desarrolló un algoritmo para detectar patrones Z y relacionados, designados como L y 7, que representan eventos de PCPA pero no se detectó la acumulación en 5' ni la disminución en 3', respectivamente (Figura 2B). Es probable que la estabilidad diferencial de los distintos transcritos producidos determine si se observa un patrón Z, 7 o L. Colectivamente, estos representan ~ 40 % de los cambios. Además, se detectaron otros patrones consistentes con PCPA, tales como todos los exones abajo (p. ej., Figura 1C; SKIV2L2), que se producen probablemente de una PCPA muy temprana, aunque no puede excluirse la posibilidad de una regulación a la baja transcripcional. Varios genes con todos los intrones arriba (p. ej., Figura 1C; NSUN2) también

tenían lecturas intrónicas poliadeniladas, indicando que estas, también, no se escapan totalmente de PCPA. Otro patrón, 3'L, que representa una clase principal a una disminución moderada de U1 como se discute más adelante, es similar a L, pero las lecturas abajo (es decir, dirección Ctrl-U1) están cerca del extremo 3' del gen, lo que es indicativo del acortamiento del transcrito. Conjuntamente, estos patrones representan el 75-90 % de los cambios en el transcriptoma en las células con U1 deplecionado, y su similitud notable en los tres organismos indica que PCPA y la función de U1 en su supresión son esenciales para la formación de transcritos de longitud completa para la mayoría de los genes de metazoos.

Una disminución moderada de U1 incita el acortamiento de 3'UTR y un cambio a isoformas de ARNm más cortas

Los datos de secuenciación de la depleción completa de U1 indicaron que PCPA ocurría ~ 10-30 nt en 3' de PAS crípticos que son similares a los PAS encontrados en el extremo 3' de los transcritos (Hu et al., 2005, RNA 11:1485-1493) (Figura 2A; Figura 9, Figura 10, y Figura 11). Estos incluyeron hexámeros canónicos (AAUAAA o AUUAAA) y raros o no descritos previamente (Beaudoing et al., 2000, Genome Res 10:1001-1010; Tian et al., 2005, Nucleic Acids Res 33:201-212), así como elementos ricos en U y ricos en GU que rodean típicamente los PAS. Para identificar los PAS que pueden ser particularmente vulnerables a PCPA y evaluar la relevancia de la supresión por U1 bajo condiciones menos drásticas, se disminuyeron los niveles de U1 un ~ 25 % y 50 %, con 0,25 y 1,0 nmoles de AMO de U1, respectivamente (Kaida et al., 2010, Nature 468:664-668). Las disminuciones moderadas de U1 se diferenciaron de las observadas en la depleción de U1 de tres maneras principales (Figura 2B). En primer lugar, se vieron afectados significativamente menos genes (~ 35-50 % en comparación con 15 nmoles; Figura 2B). En segundo lugar, no hubo una acumulación de intrones general, indicando que no se inhibió el corte y empalme (Figura 2B; p. ej. todos los intrones arriba, Z, o 7). En tercer lugar, y lo más sorprendente, la disminución moderada de U1 desplazó las posiciones de PCPA hacia distancias mucho mayores del TSS. Los ejemplos representativos de los principales patrones de distintas disminuciones en el nivel de U1 se muestran en Figura 2C. La mayoría de los cambios a la disminución moderada de U1 mostró un patrón 3'L, reflejando disminuciones cerca de los extremos de los genes (p. ej., IMPDH2) y consistieron en gran medida en disminuciones en las lecturas de 3'UTR distales (p. ej., EIF2S3, Figura 12 y Figura 13), lo que sugiere el acortamiento de 3' UTR en genes de un amplio rango de tamaños. También se observó un desplazamiento generalizado en el uso de PAS más próximas en intrones en la mitad 3' de los genes (p. ej., SHFM1). Muchos de los genes a 0,25 nmoles de AMO de U1 que tenían 3'UTR más corta (p. ej., NDUFA6), también tenían varias señales de exones reguladas al alza (Figura 2B), debido posiblemente a la eliminación de dianas de miARN en la 3' UTR (Bartel, 2009, Cell 136:215-233). El gen CD44 (Figura 2C) ilustra la PCPA continua y su dependencia de la dosis de U1, teniendo una 3'UTR más corta a 0,25, un patrón L a 1,0, y un patrón Z de PCPA en el primer intrón a 15, lo que sugiere el uso de PAS más próximos en el mismo gen con una disminución adicional de U1.

El cambio global en la posición de PCPA con el grado del nivel de la disminución de U1 se ilustra en la Figura 2D. Después de la depleción de U1 en tres organismos, la mayoría de la PCPA ocurrió ~ 1 kb del TSS, típicamente en el primer (48 %) o uno de los primeros intrones (intrón 2: 26 %; intrón 3: 11 %). Sin embargo, a menores disminuciones de U1, PCPA ocurrió a distancias mucho mayores del TSS (~ 20 kb) (Figura 2D-panel inferior, Figura 12, Tabla 2). En la mayor parte, si no todos los casos, incluyendo transcritos que terminaron en ~ 1 kb del TSS, varios PAS fuertemente predichos, y muchos más PAS no canónicos fueron puenteados antes del sitio de PCPA. De forma importante, muchos de los sitios de PCPA coincidieron precisamente con EST poliadenilados reportados previamente de un amplio rango de tipos celulares y tejidos (Figura 9 y Figura 12), lo que sugiere que PCPA es un fenómeno natural que ocurre en condiciones fisiológicas normales.

La pérdida de la supresión de PAS por U1 promueve el corte y empalme alternativo de los exones 3' terminales proximales

Se detectaron dos tipos de cambios en el ARNm que resultan de PCPA a AMO de U1 bajo (Humano 1,0 y 0,25), además de acortamiento de 3'UTR (Figura 13). En primer lugar, PCPA en intrones produjo ARNm más cortos que carecían de los exones en 3' y 3'UTR del transcrito de longitud completa. En algunos de estos, el marco de lectura abierto (ORF) del exón terminal alternativo podría extenderse potencialmente en el intrón (Figura 3A), un escenario referido como exón "compuesto" o "mezclado" (Tian et al., 2007, Genome Res 17:156-165). La RT-PCR cuantitativa en el ARN de las células transfectadas con un rango de dosis de AMO de U1 confirmó el desplazamiento a poliadenilación proximal predicho por HIDE-seq (Figura 3A). Como se esperaba, después de la depleción de U1, disminuyeron tanto las isoformas cortas como largas, debido a PCPA temprana o inhibición del corte y empalme, indicado por los patrones Z (p. ej., GABPB1 2,5 y 15 nmoles). A AMO de U1 bajo, sin embargo, en lugar de una disminución en el nivel de transcrito global, las cantidades de las formas cortas permaneció relativamente estable, mientras las formas largas disminuyeron, causando que se incrementara la relación de la isoforma corta a la larga (Figura 3A).

Un segundo tipo de lectura poliadenilada encontrada en intrones del transcrito canónico a AMO de U1 bajo (flechas sólidas) se mapeó en los extremos de exones que pueden cortarse y empalmarse alternativamente mutuamente exclusivamente con el exón terminal del gen de longitud completa ("corto" en la Figura 3), un ejemplo de APA

dependiente de corte y empalme (Edwards-Gilbert et al., 1997, *Nucleic Acids Res* 25:2547-2561; Zhang et al., 2005, *Genome Biol* 6(12):R100). El gen UBAP2L (Figura 3B) presentó un incremento dependiente de la dosis en la cantidad relativa de la isoforma más corta a AMO de U1 bajo (inserto del gel de RT-PCR). De forma interesante, se han reportado muchos casos de cambio de exón en 3, incluyendo UBAP2L, después de la activación de las células inmunes y se sugiere que resultan del corte y empalme alternativo (Sandberg et al., 2008, *Science* 320:1643-1647). Sin embargo, se consideró un mecanismo alternativo, en el que PCPA ocurre en primer lugar y que es este evento el que causa que el sitio de corte y empalme en 5' corte y empalme en un sitio de corte y empalme alternativo en 3' no utilizado en el transcrito de longitud completa. Para abordar esto, se usaron AMO para bloquear bien el sitio de corte y empalme en 3' o el PAS de la isoforma corta y se midieron los niveles de ambas isoformas por RT-qPCR (Figura 3B). Como se esperaba, el AMO frente al sitio de corte y empalme en 3' causó un cambio a la isoforma larga mediante el bloqueo del corte y empalme alternativo. De forma interesante, el AMO frente a PAS también causó que el nivel de la isoforma corta cayera drásticamente, mientras la isoforma larga se duplicó (Figura 3B; AMO de PAS). Tomados conjuntamente, estos datos sugieren que la pérdida de supresión de un PAS por U1 determina el corte y empalme alternativo a un exón terminal mutuamente exclusivo.

PCPA y su supresión es un proceso direccional 5'-3'

HIDE-seq proporcionó información acerca de la posición de PCPA en intrones, revelando que ocurría en los tres organismos a una distancia media de 500-1.000 nt del sitio de corte y empalme en 5' más cercano (Figura 4A), pero a distancias variables y habitualmente mucho mayores del sitio de corte y empalme en 3' más cercano (Figura 4A) (Figura 9, Figura 10, Figura 11, y Tabla 2). En cada organismo, se encontraron casos de múltiples sitios PCPA en el mismo intrón (p. ej., Spen, RBM39) o mismo transcrito (p. ej., Spen, S1c38a2), demostrando que hay varios PAS accionables a lo largo de un transcrito (Figura 4B) y es probable que se requiera U1 unido en el sitio de corte y empalme en 5' y en otra parte en intrones para suprimirlos.

Para ensayar el mecanismo de PCPA, se investigó si un U1 defectivo en corte y empalme puede suprimir PCPA, usando construcciones de minigen NR3C1 (Figura 4C) (Kaida et al., 2010, *Nature* 468:664-668). El minigen WT (carril 1) corta y empalma apropiadamente, mientras una mutación en el sitio de corte y empalme en 5' causa PCPA 385 nt en el intrón 2 (carril 2). PCPA se suprimió completamente por un U1 mutado en el extremo 5' complementario al sitio de corte y empalme en 5' (mutU1/B). Se observaron grados de supresión variados también para U1 unidos en otros lugares, en el intrón y exón en 5' ambos en la proximidad del PAS críptico (~ 45-80 %), e incrementando U1 WT. Esto sugiere que U1 puede suprimir PCPA incluso sin ser capaz de funcionar en el corte y empalme y que U1 unido en 5' lo hace más eficientemente (p. ej., mutU1/A y mutU1/C). A continuación, se construyó un minigen NR3C1 en el que el PAS accionable en el intrón 2 se duplicó (Figura 4D). La transfección de esta construcción con AMO de U1 (carril 2) causó PCPA en el primer PAS (PAS1; 385 nt) y como se ha reportado previamente, una mutación en el sitio de corte y empalme en 5' causó PCPA (carril 3), pero ocurrió más a partir de este PAS en presencia de AMO de U1 (carril 4) (Kaida et al., 2010, *Nature* 468:664-668). Cuando se mutó PAS1, PCPA ocurrió ahora a partir del segundo PAS (PAS2) localizado 1295 nt del sitio de corte y empalme en 5', indicando que un PAS(s) en 3' es también vulnerable y que PCPA ocurre con una direccionalidad 5' a 3'. Cuando se mutaron tanto el sitio de corte y empalme en 5' como PAS1, ocurrió algo de PCPA a partir de PAS2 (carril 7), indicando que la protección desde el sitio de corte y empalme en 5' se extiende fuera a PAS2, pero, de nuevo, el AMO de U1 incitó más PCPA (carril 8), lo que sugiere que U1 proporciona una supresión adicional de PCPA unido a secuencias distintas del sitio de corte y empalme en 5'.

Para abordar cuánto se extiende la supresión de PCPA por la RNPs U1 más allá de un sitio de corte y empalme en 5', se estudió el gen endógeno BASP1, que se somete a PCPA relativamente lejos del sitio de corte y empalme en 5', ~ 3,5 kb en el primer intrón (Figura 4E). De forma interesante, poco o nada de PCPA ocurrió con un AMO bloqueante del sitio de corte y empalme en 5' (carril 3), mientras la cotransfección con el AMO frente al sitio de corte y empalme en 5' y AMO de U1 (carril 4) dio lugar a una cantidad sustancial, lo que sugiere que la supresión por U1 del sitio de corte y empalme en 5' solo es insuficiente. Los controles adicionales confirmaron que el AMO frente al sitio de corte y empalme en 5' era efectivo en la inhibición del corte y empalme (carriles 2-4: ARNm disminuye e IR1 se incrementa) y no se activó un sitio de corte y empalme en 5' críptico (Figura 4E). Se concluye que, a una distancia de 3,5 kb, un PAS está fuera del rango protector de U1 unido al sitio de corte y empalme en 5'. Tomados conjuntamente, los datos de HIDE-seq y el ensayo directo demuestran la direccionalidad 5' a 3' del proceso PCPA y sugieren fuertemente una necesidad de U1 en exceso de lo que se requiere para el corte y empalme.

Tabla 2: resumen de las distancias de PCPA

Se calcularon la mediana, media y desviación estándar para los valores presentados en la Figura 9, Figura 10, Figura 11, y Figura 12.

	Humano (15)			Ratón (15)			Drosophila (15)			Humano (1.0/25)		
	Mediana	Media	sd	Mediana	Media	de	Mediana	Media	de	Mediana	Media	de
Longitud génica	38780	83009	132728	24827	52568	76670	11006	20305	23246	37679	59771	57869
Distancia desde TSS	3413	20175	69988	1445	6491	14333	1639	4075	7570	15359	25164	31914
Distancia desde TSS más cercano	1965	8269	16591	1306	2391	3045	1377	3104	7269	nd	nd	nd
Número de Intrones	2	2.1	1.7	1	2.1	1.8	2	2.4	1.9	5	7.9	10.4
Longitud de Intrones	7340	25844	46182	3780	17793	40755	2091	8059	14830	NA	NA	NA
Distancia desde sitio de corte, y empalme en 5'	570	800	1368	549	741	655	346	555	878	NA	NA	NA
Distancia desde sitio de corte, y empalme en 3'	6752	25044	45340	3190	17052	40531	1546	7510	14785	NA	NA	NA

La disminución de U1 recapitula el acortamiento de isoformas en neuronas activadas

Los ejemplos consistentes con el acortamiento de transcritos descritos en la presente memoria se han caracterizado en neuronas activadas, particularmente el gen *homer-1/vesl-1*, que juega un papel crítico en la sinaptogénesis. La activación neuronal da lugar a un desplazamiento rápido (< 6 hr) en el procesamiento del pre-ARNm que codifica *homer-1*, de un ARNm de longitud completa (L) a uno más corto (S). *Homer-S* se produce por una extensión del exón 5 y APA en el intrón en 3' para delecionar los exones que codifican el C-terminal (Niibori et al., 2007, Neurosci Res 57:399-410). Usando células PC12 de rata, se muestra por RT-PCR que un rango de AMO de U1 bajo (0,25-0,5 nmoles) incrementó dramáticamente la forma S y causó una disminución recíproca dependiente de la dosis de *homer-L*, de una forma idéntica al cambio observado después de la activación neuronal con forskolina y KCl (Figura 5A). Este desplazamiento en la relación de S/L (histograma) fue evidente incluso a 0,1 nmoles, correspondiente a una disminución estimada del 10 % en U1. La disminución adicional de U1 (1,0-4,0 nmoles) causó que ambas formas desaparecieran, probablemente como resultado de un desplazamiento de PCPA más cercana al TSS. Además de *homer-1*, se ha descrito el acortamiento dependiente de actividad de muchas otras proteínas críticas para la sinaptogénesis, incluyendo *Dab1* (Flavell et al., 2008, Neuron 60:1022-1038). De hecho, en células MN-1 de ratón, *Dab1* también presentó un desplazamiento similar de isoformas con una disminución de U1 (Figura 5B). Los resultados presentados en la presente memoria sugieren que los niveles de U1 pueden regular el tamaño de las isoformas de ARNm producidas a partir de muchos genes.

La regulación al alza transcripcional en neuronas activadas crea una escasez de U1

Dos posibles escenarios podrían explicar potencialmente el desplazamiento en el uso de PAS más proximales en las células activadas, si esto se debiera de hecho a la pérdida de la supresión de PCPA por U1. Bien los niveles de U1 disminuyeron o se incrementó la cantidad de transcritos nacientes que es necesario que proteja. Para explorar estas posibilidades en condiciones fisiológicas, se determinaron la cantidad de U1 y pre-ARNm durante la activación neuronal (Figura 6A). Mediante la comparación de células PC12 control y activadas, U1 se cuantificó por RT-qPCR, y los transcritos nacientes se marcaron por pulso con 5'6'-³H uridina durante 30 min antes del aislamiento del ARN y selección del ARN poli(A), seguido de conteo por centelleo para determinar la relación de ARNm naciente frente a ARN total. Esto reveló un incremento robusto en el resultado transcripcional de aproximadamente el 40-50 % a las 2-4 hrs después de la activación, mientras los niveles de U1 no cambiaron, creando una escasez de U1 significativa respecto a sus dianas. Este hueco volvió a la línea base a las 8 hrs, proporcionando una ventana integrada de oportunidad para una rápida regulación de la expresión génica global en respuesta a estímulos externos. Notablemente, las cantidades de *Homer-S* coincidieron con la ventana de escasez de U1, elevándose mucho después de la activación, formando un pico a las 2-4 horas (incremento de ~ 10 veces) y volviendo a niveles cercanos a la línea base a las 6-8 horas. Por el contrario, la forma predominante, *Homer-L*, permanece casi inalterada después de la activación.

La sobreexpresión de U1 inhibe el acortamiento de ARNm en neuronas activadas

Mientras la escasez de U1 crea una oportunidad para usar los PAS proximales, el cambio a isoformas más cortas puede no resultar necesariamente de la pérdida de la función de U1. Para abordar esto, se examinó si la sobreexpresión de U1 podría contrarrestar el cambio de *Homer-L* a *Homer-S* en las neuronas activadas. Las células PC12 se transfectaron con cantidades crecientes de un vector de expresión de ARNs de U1 o un vector vacío 24 hr antes de la activación. La U1 expresada endógenamente, cuantificada por RT-qPCR, mostró que la sobreexpresión (~ 40% de U1 endógeno) evitaba el cambio de *homer-1* a la isoforma *Homer-S* de una manera dependiente de la dosis (Figura 6B). Esto sugiere un papel para la supresión de PCPA por U1 en la ruta de activación neuronal y demuestra que la alteración de los niveles de U1 regula la expresión génica.

Telescripting de U1

Las matrices en mosaico genómicas parciales mostraron previamente que la protección de U1 es necesaria para prevenir la terminación prematura drástica de la mayoría de los transcritos pol II nacientes por PCPA de PAS crípticos diseminados a lo largo de los intrones (Kaida et al., 2010, Nature 468:664-668). Esta actividad se refiere como *telescripting*, ya que es necesaria para que los transcritos nacientes se extiendan sobre largas distancias. Aquí, una estrategia de alto rendimiento rápida y versátil para identificar cambios en el transcriptoma, HIDE-seq, y experimentos basados en la abundante información que proporciona, rindió una mayor definición del *telescripting* de U1, revelando que no es solo una medida conservada evolutivamente para asegurar la integridad del transcriptoma, sino también un mecanismo robusto para la regulación de la expresión génica. Sorprendentemente, la posición de PCPA en un gene dado varió dependiendo de la cantidad de U1 disponible. Aunque la depleción de U1 inhibió el corte y empalme y causó PCPA, típicamente en el primer intrón, las disminuciones moderadas de U1 (10-50 %) no inhibieron el corte y empalme, sino que causaron PCPA más en 3', tendiendo hacia mayores distancias desde el TSS con menores disminuciones de U1. Significativamente, se vieron afectados menos genes, pero todavía numerosos, por una disminución moderada de U1 en comparación con la depleción de U1 y no fue destructora, produciendo ARNm más cortos debido al uso de PAS alternativos, más próximos al promotor, en lugar del PAS normal en el extremo 3' del gen de longitud completa. Así, el *telescripting* puede modularse disminuyendo el U1 disponible en un gran rango sin comprometer el corte y empalme, consistente con la idea de que hay un exceso de U1 por encima de lo que se requiere para el corte y empalme. En la presente memoria se describe que el *telescripting* proporciona un mecanismo de regulación de la expresión génica global.

Los datos presentados en la presente memoria demuestran que los cambios predominantes en el transcriptoma que resultan del *telescripting* incompleto son formas variadas de acortamiento de ARNm, incluyendo acortamiento de 3'UTR y PCPA en intrones. Así, el *telescripting* juega un papel importante en la determinación de la longitud del ARNm y la expresión de sus isoformas. De hecho, pueden producirse varios ARNm con diferentes longitudes a partir del mismo gen por PCPA, correspondiente al grado de disminución de U1 (p. ej., Figura 3, GABPB1, UBAP2L). De forma importante, PCPA puede modular profundamente los niveles de expresión e isoformas de las proteínas. Aunque el acortamiento de 3' UTR no cambiaría la secuencia de la proteína codificada, elimina elementos, tales como sitios de unión de microARN y proteína RNPhn, que podrían ser críticos para la regulación del ARNm, incluyendo su estabilidad, localización, y eficiencia en la traducción (Filipowicz et al., 2008, Nat Rev Genet 9:102-114; Huntzinger e Izaurralde, 2011, Nat Rev Genet 12:99-110). Por el contrario, PCPA en intrones produciría un ARNm que codifica una proteína que carece del extremo C o que contiene un nuevo péptido C-terminal, si el ORF se extiende en el intrón terminal. Un escenario adicional asociado con PCPA intrónica es el cambio de exones en 3', también referido como APA dependiente de corte y empalme, que refleja la visión general de que mecánicamente, el corte y empalme alternativo determina el PAS que se utiliza. Sin embargo, en la presente memoria se mostró que el enmascaramiento por AMO del PAS del exón terminal alternativo previno su corte y empalme, indicando que el corte y empalme en el exón terminal alternativo dependía de PCPA de este PAS (Figura 3B). Así, PCPA puede ser el evento primario en el cambio de exones en 3', revelando un papel inesperado independiente del corte y empalme para U1 en la regulación del corte y empalme alternativo.

Varias líneas de evidencia sugieren fuertemente que el *telescripting* de U1 es un fenómeno fisiológico. En primer lugar, en el rango completo de AMO de U1 usado, numerosos sitios de PCPA coincidieron precisamente con transcritos poliadenilados detectados previamente, indicando que estos sitios crípticos se utilizan naturalmente. Estos incluyen EST que representan isoformas de ARNm cortas de un amplio rango de espécimen (Figura 9 y Figura 12), de las cuales algunas se han indicado como APA de PAS proximales (Lou et al., 1996, Genes Dev 10:208-219; Pan et al., 2006, Gene 366:325-334; Tian et al., 2007, Genome Res 17:156-165). En segundo lugar, muchos de los eventos de acortamiento de ARNm que resultan de la pérdida de supresión de PCPA son indistinguibles del acortamiento de ARNm general observado en linfocitos T y neuronas activadas, células proliferantes y células cancerosas (Flavell et al., 2008, Neuron 60:1022-1038; Mayr y Bartel, 2009, Cell 138:673-684; Sandberg et al., 2008, Science 320:1643-1647; Zhang et al., 2005, Genome Biol 6(12):R100). Notablemente, el ~ 33 % de los genes que experimentan acortamiento de 3' UTR en las células T activadas (Sandberg et al., 2008, Science 320:1643-1647) se ven afectados de manera similar por AMO de U1 bajo en células HeLa, como RAB10 y CCND1 (Figura 2C y Figura 13), observado en células cancerosas (Mayr y Bartel, 2009, Cell 138:673-684). En tercer lugar, una disminución moderada de U1 recapituló precisamente los cambios característicos a isoformas más cortas en un modelo de activación neuronal. Usando inductores nativos (KCI y forskolina), se muestra que la disminución de U1 sola causa el mismo cambio de isoformas tanto en *homer-1* como *Dab-1* de una manera dependiente de la dosis. Representando el ejemplo mejor caracterizado, el gen *homer-1* cambia a una isoforma que carece de los exones que codifican el dominio C-terminal, que antagoniza la actividad crítica de la proteína de longitud completa en el reforzamiento de las sinapsis y potenciación a largo plazo (Sala et al., 2003, J Neurosci 23:6327-6337).

Habiendo basado estos estudios en la regulación a la baja deliberada de U1, se consideró si hay circunstancias fisiológicas bajo las cuales los niveles de U1 podrían volverse deficientes. Dada la abundancia y muy larga semivida de U1 (Sauterer et al., 1988, Exp Cell Res 176: 344-359), parecía improbable que sus niveles disminuirían significativamente en términos absolutos en el marco de tiempo corto en el que se observa el acortamiento de ARNm.

Se consideró un escenario alternativo mediante el cual la escasez de U1 respecto a las dianas que es necesario que proteja en transcritos nacientes podría surgir simplemente por un incremento en el rendimiento transcripcional de los pre-ARNm. Apoyando este escenario, las mediciones mostraron un incremento rápido y transitorio en los transcritos nacientes de ~ 40-50 % después de la activación neuronal, mientras los niveles de U1 mostraron un cambio pequeño, si existió (Figura 6A). Esto crea una escasez significativa de U1 respecto a los pre-ARNm nacientes, cuya magnitud está en el mismo rango de los experimentos con AMO. Este hueco dirigido por la transcripción, detectable a las 2-4 hr y que vuelve a la línea base a las 6-8 hr después de la activación, crea una ventana de oportunidad para que ocurra APA de PAS proximales debido a la disminución transitoria en la capacidad de *telescripting*. De forma importante, el cambio a isoformas más cortas durante la activación neuronal podría verse antagonizado por la sobreexpresión de U1 de una manera dependiente de la dosis. Estos datos proporcionan una evidencia adicional de que la supresión de PCPA por U1 es un componente en el mecanismo de selección de PAS, y juega así un papel importante en la regulación de la expresión génica durante la activación neuronal y potencialmente en respuesta a otros estímulos fisiológicos.

El mecanismo y el o los factores implicados en el acortamiento de ARNm en diversas condiciones de activación no se han identificado (Flavell et al., 2008, Neuron 60:1022-1038; Ji y Tian, 2009, PLoS One 4:e8419; Mayr y Bartel, 2009, Cell 138:673-684; Sandberg et al., 2008, Science 320:1643-1647). Sin embargo, las similitudes notables que tienen con la escasez de U1 sugieren que también implican la pérdida del *telescripting* de U1, al menos durante la fase inicial (inmediata/temprana). Se han descrito otros factores que también podrían causar un desplazamiento a PAS alternativos proximales, particularmente la regulación al alza de la poliadenilación general y factores de procesamiento del extremo 3', tales como Cstf64 (Chuvpilo et al., 1999, Immunity 10:261-269; Takagaki et al., 1996, Cell 87:941-952). Sin embargo, esto requiere nueva síntesis de proteínas y tarda muchas (>18) horas (Shell et al., 2005, J Biol Chem 280:39950-39961), y, por lo tanto, no puede explicar el cambio rápido a PAS proximal, a diferencia de la escasez de U1, que es inmediata después de la estimulación y ocurre incluso en presencia de la inhibición de la síntesis de proteínas (Loeblich y Nedivi, 2009, Physiol Rev 89(4):1079-1103). El papel potencial de U1 y otros factores posteriormente después de la activación celular y en otras células, permanece sin determinar. Otros medios para crear escasez de U1, sin incremento en la transcripción, pueden concebirse, tales como su secuestro en estructuras nucleares o por la expresión de otros ARN a los que podría unirse.

Se propone un modelo de *telescripting* de U1 que podría explicar su papel en la regulación de la longitud del ARNm y en el cambio de isoformas (Figura 7). Aunque no se pretende la vinculación a ninguna teoría particular, se sugiere que PCPA ocurre cotranscripcionalmente por la misma maquinaria que lleva a cabo la escisión y poliadenilación normal del extremo 3' en el exón terminal del gen de longitud completa, y es un subproducto del acoplamiento entre la transcripción y este proceso aguas abajo (Calvo y Manley, 2003, Genes Dev 17:1321-1327; Dantonel et al., 1997, Nature 389:399-402; McCracken et al., 1997, Nature 385:357-361). Los factores del CPA se asocian con el TEC de la pol II (Glover-Cutter et al., 2008, Nat Struct Mol Biol 15:71-78; Hirose y Manley, 1998, Nature 395:93-96) cerca del TSS y están, por lo tanto, listos para procesar PAS recién transcritos con secuencia y características estructurales favorables (PAS accionables) que se encuentra a lo largo de la mayor parte de la longitud del gen. Esto se evita normalmente por la RNPsn U1 que se une al transcrito naciente. Los estudios previos han mostrado que U1 puede inhibir la poliadenilación del PAS normal cuando se une en su proximidad en el exón terminal e *in vitro* (Ashe et al., 2000, RNA 6(2):170-177; Fortes et al., 2003, Proc Natl Acad Sci USA 100:8264-8269; Gunderson et al., 1998, Mol Cell 1:255-264; WO 2008/121963; Vagner et al., 2000, RNA 6(2):178-188). U1 se recluta en transcritos pol II nacientes, incluyendo transcritos sin intrones, por múltiples interacciones con el pre-ARNm y factores de procesamiento del ARN, así como la maquinaria transcripcional (Brody et al., 2011, PLoS Biol 9:e1000573; Das et al., 2007, Mol Cell 26:867-881; Lewis et al., 1996, Genes Dev 10:1683-1698; Lutz et al., 1996, Genes Dev 10:325-337). En los tres organismos estudiados, PCPA no se detectó típicamente en los primeros pocos cientos de nucleótidos, elevándose mucho posteriormente y alcanzando el pico ~ 1 kb del TSS (Figura 2D). Es posible que los PAS accionables en 5' de este punto no se utilicen porque los factores del CPA pueden no estar todavía asociados con el TEC (Mayer et al., 2010, Nat Struct Mol Biol 17:1272-1278; Mueller et al., 2004, Mol Cell 14(4):447-456), que depende del estado de fosforilación del dominio carboxilo terminal de pol II (Buratowski, 2009, Mol Cell 36:541-546). Aunque no se pretende la vinculación a ninguna teoría particular, se sugiere que este retraso podría servir (y puede haber evolucionado) para permitir la unión de U1 antes de que los transcritos se expongan al CPA, para evitar su destrucción temprana. Consistente con esta consideración de la distancia desde el TSS, los PAS en la 5'UTR que no han experimentado PCPA son funcionales cuando se ponen en el extremo 3' del gen (Guo et al., 2011, Mol Cell Biol 31:639-651).

Los datos presentados en la presente memoria también indicaron que el U1 unido al sitio de corte y empalme en 5' tiene un rango protector limitado de hasta ~ 500-1.000 nt y, por lo tanto, sería insuficiente para asegurar el *telescripting* a través de intrones mayores. La supresión completa de PCPA no podría proporcionarse por U1 desde el sitio de corte y empalme en 5' solo incluso en este perímetro y no juega casi ningún papel en la protección de PAS más distales, dependiendo, en su lugar, de U1 unido adicional en intrones (Figuras. 4D y 4E). De hecho, los intrones contienen numerosos sitios de unión de U1 que no funcionan como sitio de corte y empalme en 5' que, sin pretender la vinculación a ninguna teoría particular, se sugiere que sirven para anclar U1 para proteger a los intrones frente a PCPA. Además,

se encontró que U1 con una secuencia 5' mutada que no puede funcionar en el corte y empalme, todavía puede funcionar en *telescripting* (Figura 4C), apoyando la noción de dos papeles separados de U1. La capacidad de U1 de interactuar con transcritos nacientes directamente, incluso sin emparejamiento de bases (Patel et al., 2007, J Cell Biol 178:937-949; Spiluttini et al., 2010, J Cell Sci 123:2085-2093), aumenta su asociación a lo largo del pre-ARNm, permitiéndole barrer el transcrito y acelerar la tasa mediante la que puede encontrar sitios de emparejamiento de bases más estables, incluyendo el sitio de corte y empalme en 5'. De forma importante, las mutaciones en PAS causan PCPA de PAS intrónicos en 3' (Figura 4D), lo que sugiere un proceso direccional, consistente con un mecanismo cotranscripcional. Se propone que la escasez de U1 causa el acortamiento del transcrito al volverse limitante el reclutamiento cotranscripcional de U1 a transcritos nacientes, deja a los PAS distales menos protegidos, proporcionando un componente y un mecanismo dependiente de la dosis de U1 para la regulación de la longitud del ARNm y cambio de isoformas.

Ejemplo 2: la sobreexpresión del ARNsn de U1 revierte el acortamiento de 3'UTR de las células cancerosas y modera el fenotipo agresivo

El acortamiento de ARNm general, que resulta de la poliadenilación alternativa (APA) en señales de poliadenilación crípticas (PAS) más proximales en las regiones no traducidas en 3' (UTR) y en intrones se ha asociado con la proliferación celular, cáncer, activación de células inmunes y neuronas, y crecimiento de células madre (Flavell et al., 2008, Neuron 60(6):1022-38; Ji et al., 2009, PLoS One 4(12):e8419; Mayr y Bartel, 2009, Cell 138(4):673-84; Sandberg et al., 2008, Science 320(5883):1643-7). Este fenómeno llamativo ha recibido una atención particular porque el acortamiento de 3'UTR elimina frecuentemente los sitios de unión de microARN (miARN) enriquecidos en esta parte de los ARNm, dando lugar a la pérdida de la regulación de los miARN. Esto alivia típicamente la represión de la traducción mediada por los miARN dando lugar frecuentemente a un fuerte incremento (>10 veces) en la cantidad de proteína producida a partir del ARNm acortado, incluyendo proto-oncogenes (Mayr y Bartel, 2009, Cell 138(4):673-84; Filipowicz et al., 2008, Nat Rev Genet 9(2):102-14). Los estudios recientes han identificado el acortamiento de 3'UTR en >200 genes como una firma de cáncer agresivo de pulmón y de mama con mal pronóstico (Lembo et al., 2012, PLoS One 7(2):e31129). A pesar de la importancia de este fenómeno, son desconocidos el o los factores que median el acortamiento de ARNm en el cáncer, células proliferantes y bajo diversas fisiologías de activación.

La RNPsn U1 (U1) es un complejo de ribonucleoproteínas (RNP) abundante comprendido por el ARNsn de U1 no codificante (165 nt en mamíferos) y proteínas de unión a ARN específicas que funcionan en el reconocimiento del sitio de corte y empalme en 5' (ss), la primera etapa en el corte y empalme (Wahl et al., 2009, Cell 136(4):701-18). En estudios recientes, se observó que la depleción de la RNPsn U1 causaba la escisión y poliadenilación prematuras (PCPA) en la mayoría de los pre-ARNm en PAS crípticos encontrados a lo largo de los transcritos nacientes (Kaida et al., 2010, Nature 468(7324):664-8). Mediante la supresión de PCPA, U1 permite que los pre-ARNm se extiendan más de manera que se usan señales de poliadenilación (PAS) más distales (Berg et al., 2012, Cell 150(1):53-64). Esta actividad se denomina *telescripting*, y se mostró que es específica de U1 para este ARNsn y que está separada de su función en el corte y empalme. El *telescripting* efectivo depende de la abundancia de U1, en exceso de lo que se necesita para el corte y empalme, e incluso una disminución moderada de U1 respecto a las dianas ARN nacientes, por ejemplo, como ocurriría debido a la regulación al alza de la transcripción, podría incitar el acortamiento del ARNm, incluyendo en las 3'UTR (Berg et al., 2012, Cell 150(1):53-64). Así, los cambios en la actividad de *telescripting* podrían proporcionar un mecanismo de APA, que es desconocido en gran medida (Di Giammartino et al., 2011, Mol Cell 43(6):853-66), así como explicar el acortamiento de 3'UTR general en el cáncer.

En los experimentos presentados en la presente memoria, los niveles de U1 se modularon en un sistema de células cancerosas definido y se estudiaron sus consecuencias para la longitud de las 3'UTR y el fenotipo del cáncer. En la presente memoria se muestra que la sobreexpresión de U1 alargó numerosas (>2.100) 3'UTR ya acortadas en células cancerosas HeLa. A la inversa, la disminución de U1 acortó 3'UTR adicionales (>700). De forma notable, el incremento de U1 atenuó (~ 50 %) la migración y capacidad invasiva celulares, mientras su disminución las aumentó de manera dependiente de la dosis (~ 500 %). Muchos de los cambios en las longitudes de las 3'UTR recapitularon la desregulación del miARN diana causante del cáncer en oncogenes. Los datos presentados en la presente memoria demuestran que el acortamiento del ARNm en el cáncer puede explicarse por déficit de *telescripting* y que su modulación por U1 solo podría modificar profundamente y bidireccionalmente la agresividad del cáncer, indicando de esta manera que el *telescripting* es una diana para moderar la tumorigénesis.

Se describen ahora los materiales y métodos empleados en este ejemplo experimental.

Material y métodos:

Cultivo celular, oligonucleótido morfolino antisentido (AMO), y sobreexpresión de U1

Las células de mantuvieron en DMEM (HeLa, A549, MCF-7 y MB-231) suplementado con suero fetal bovino (FBS) al 10 %, 10 unidades/ml de penicilina y 10 µg/ml de estreptomycin a 37 °C. y CO₂ al 5 %. Las secuencias de los AMO

de U1 y U2 y las transfecciones fueron como se ha descrito previamente (Kaida et al., 2010, Nature 468(7324):664-8). La sobreexpresión de U1 se consiguió transfectando células HeLa durante 24 hr con un plásmido de expresión de U1 (1 y 1,5 µg) en el vector pSiren-RetroQ (Clontech) usando el reactivo Effectene (Qiagen) según las instrucciones del fabricante. El promotor U6 se reemplaza por el promotor de U1 nativo para dirigir la expresión de U1. Se construyó un vector vacío con el mismo promotor de U1 para usarse como un control.

Ensayo de proliferación

La proliferación celular se midió usando el kit del ensayo de viabilidad celular luminiscente CellTiter-Glo (Promega) según las instrucciones del fabricante. Las células se transfectaron con los AMO o sobreexpresión de U1 y se sembraron en triplicado en placas de 96 pocillos a una densidad de 1×10^4 células por pocillo. Las células se incubaron en medio que contenía FBS al 1 % durante 3 días y la proliferación se midió cada 24 horas.

Ensayos de migración e invasión

La migración se evaluó con el ensayo de migración celular de 24 pocillos Cytoselect (Cell Biolabs) y la invasión se midió usando cámaras de invasión de Matrigel BD BioCoat (BD Bioscience). Brevemente, en ambos ensayos, las células se transfectaron con los AMO o plásmido que expresaba U1, y se sembraron 5×10^5 células/ml por pocillo en una cámara superior en medio sin suero. La cámara inferior se llenó con medio que contenía FBS al 10 %. Después de 24 hr, las células de la cámara inferior se tiñeron y se contaron según las instrucciones del fabricante.

Marcaje metabólico del ARN, aislamiento y RNA-Seq

Las células se marcaron durante 2 hrs con 4-tiouridina 250 µM añadida a las células entre 6-8 hr después de la transfección con AMO de U1 o entre 22-24 hrs después de la transfección con plásmidos que expresaban U1 o control. Se extrajo el ARN total con Trizol (Invitrogen) y el ARN poli(A) se purificó en lechos Oligotex (Qiagen). Se hicieron reaccionar los tioles libres del ARNm poli(A) con 0,2 mg/ml de biotina-HPDP durante 2 hr para marcar el ARN incorporando 4-tiouridina. El ARN se purificó entonces en Dynabeads con estreptavidina M-280 (Invitrogen), el ADNc se sintetizó usando un kit sistema Ovation RNA-Seq V2 (NuGEN) y se construyeron bibliotecas para secuenciación por Illumina usando el sistema de biblioteca Encore NGS según las instrucciones del fabricante.

Mapeo de las lecturas de RNA-Seq

Las lecturas de RNA-seq se alinearon con el genoma de referencia UCSC/hg19 usando Tophat (versión 1.3.1) (Trapnell et al., 2009, Bioinformatics 25(9):1105-11) con ajustes por defecto. Las lecturas por exón se agruparon, de las que se calcularon los valores RPKM (Lecturas por kilobase de modelo de exón por millón de lecturas mapeadas (Mortazavi et al., 2008, Nat Métodos 5(7):621-8)). Solo se incluyeron genes con RPKM>1 en los análisis adicionales.

Análisis de la expresión génica diferencial

Se usó la herramienta Cuffdiff de Cufflinks (Trapnell et al., 2010, Nature Biotechnology 28(5):511-5) (versión 1.3.0) para identificar a los genes expresados diferencialmente significativamente, comparando la relación log de la expresión de un gen en dos condiciones frente al log de uno. Los genes afectados significativamente se reportaron con una FDR de 0,05 usando la corrección de Benjamini-Hochberg para ensayos múltiples.

Análisis de corte y empalme alternativo

El análisis MISO (versión 0.4.1) (Katz et al., 2010, Nat Methods 7(12):1009-15) se realizó en datos de RNA-seq alineados para identificar a las isoformas con corte y empalme diferencial en muestras experimentales en comparación con los controles y para cuantificar sus niveles de expresión computando el PSI (porcentaje de isoforma con corte y empalme; Ψ) y factores de Bayes. Los eventos de corte y empalme alternativo se filtraron según los siguientes criterios: (a) al menos 15 lecturas de inclusión, (b) al menos 15 lecturas de exclusión, de manera que (c) la suma de lecturas de inclusión y exclusión es al menos 30, (d) el $|\Delta\Psi|$ es al menos 0,3 (e) el factor de Bayes es al menos 10, y (a)-(e) son verdaderas en una de las muestras.

Algoritmo UTRend

Para un gen dado, se extrajo la 3'UTR de su isoforma canónica más larga de genes RefSeq y se mapearon las localizaciones de todos los posibles sitios APA o extremos 3' del transcrito en esta región usando las bases de datos PolyA_DB UCSC y AceView como referencias. Para incluir nuevos extremos 3' potenciales no encontrados en ninguna de las anotaciones mencionadas anteriormente, pero presentes en el conjunto de datos, se desarrolló un algoritmo de búsqueda de picos para representar el patrón de lecturas de secuencias mapeadas de 3'UTR para un gen. Se infirió un extremo 3' potencial como sigue: 1) se calculó la distribución de cobertura de la lectura a lo largo de la 3'UTR completa en muestras control y de tratamiento separadamente; 2) las dos distribuciones se fusionaron por la función MAX; 3) los datos se suavizaron entonces por LOWESS y se buscó el máximo local como picos. Para cada extremo

3' anotado y predicho, se computaron las cuentas de lecturas en su ventana de 200 nucleótidos en 5' como la estimación del nivel de expresión de la isoforma de ARNm que terminaba en ese extremo 3'. Se calculó la relación relativa de expresión en picos proximales a distales, además de un valor p, por el ensayo exacto de Fisher usando R. Sin perder la generalidad, la predicción de alargamiento puede ser equivalente a la predicción de acortamiento en la otra dirección y, así, se usaron una relación relativa >3 y un valor p <0,01 para predecir la tendencia del cambio en la longitud de las 3'UTR.

Análisis de los sitios diana de miARN

Se descargaron los sitios reguladores de miARN de TargetScan del navegador de genoma UCSC. La coordenada de cada sitio de miARN predicho se comparó con aquellos genes para los que la 3'UTR mostró cambios en su longitud. Para las Figuras. 14 y 15, solo se muestran los miARN que se expresan en HeLa y que se ha demostrado que están implicados en el cáncer.

Marcaje metabólico de proteínas

Las células HeLa se transfectaron con AMO de U1 o U2 durante 8 horas, o con 1,5 µg de plásmido de expresión de U1 durante 24 horas. Se usó cicloheximida como control para la inhibición completa de la traducción. Tres horas antes de la recogida, las células se desprovieron de suero durante 2 hrs, se marcaron entonces con [³⁵S]metionina Promix (40 µCi/ml) durante la última hora. Las proteínas sintetizadas *de novo* se lisaron por hidrólisis alcalina, se precipitaron con ácido tricloroacético, se capturaron en papeles de filtro y se lavaron extensamente para eliminar la [³⁵S]metionina no incorporada. La radiactividad se midió por conteo de centelleo en un TriCarb 2900TR.

Se describen ahora los resultados de los experimentos.

Telescripting como una diana potencial para moderar la tumorigénesis

Para determinar si el acortamiento de 3'UTR en las células cancerosas está causado por déficit de *telescripting*, la U1 se sobreexpresó en células de carcinoma de cuello uterino humano (HeLa) y se usó una secuenciación profunda de ARN (RNA-seq) para caracterizar los cambios resultantes en la secuencia en ARN poliadenilados (ARNm y pre-ARNm). La sobreexpresión de U1 del 20 % y 40 % se consiguió transfectando 1,0 µg y 1,5 µg, respectivamente, de un plásmido que contenía el gen de ARNsn de U1 bajo el control de su promotor nativo (Figura 22). El mismo vector sin un gen de U1 se usó como control (vector vacío). Para la comparación, también se examinó el efecto de la disminución de U1. Sobre la base de estudios previos, U1 se disminuyó solo moderadamente, por transfección de dosis bajas de oligonucleótido morfolino antisentido complementario al extremo 5' del ARNsn de U1 (AMO de U1) (Kaida et al., 2010, Nature 468(7324):664-8; Berg et al., 2012, Cell 150(1):53-64), que ocluye la secuencia de 9 nucleótidos en 5' de U1 que es necesaria para sus actividades en el corte y empalme y *telescripting*. Las transfecciones con AMO de U1 (0,1-1,0 nmoles) en células HeLa inactivaron ~ 15-40 % de U1 en <8 hr (Kaida et al., 2010, Nature 468(7324):664-8). Se usó una dosis de 0,25 nmoles para el RNA-seq ya que produjo los máximos cambios en el fenotipo descritos más adelante. Como el *telescripting* ocurre cotranscripcionalmente, los cambios en la longitud del ARNm se detectarían más fácilmente en los transcritos nacientes. Así, los ARN se marcaron metabólicamente con 4-tiouridina (Dolken et al., 2008, RNA 14(9):1959-72; Rabani et al., 2011, Nat Biotechnol 29(5):436-42) durante 2 hrs, se seleccionaron y se secuenciaron solo los transcritos poliadenilados nacientes producidos en la ventana de tiempo durante la que están alterados los niveles de U1, como se describe en otro lugar en la presente memoria.

En la Tabla 3 se muestra un resumen de los datos de RNA-seq para la sobreexpresión de U1 y AMO de U1, en comparación con sus controles respectivos. Se obtuvieron entre 95 y 226 millones de lecturas de extremo emparejado, de las cuales ~ 50-60 % pudieron mapearse únicamente en el genoma de referencia humano (UCSC, hg19) usando TopHat (Trapnell et al., 2009, Bioinformatics 25(9):1105-11). Las lecturas se agruparon por genes anotados RefSeq y solo aquellas para las que una profundidad de secuenciación suficiente, calculada a partir del número de lecturas por kilobase por millón de lecturas mapeadas (Mortazavi et al., 2008, Nat Métodos 5(7):621-8) (RPKM)>1, se incluyeron en análisis adicionales. Esto consistió en 12.639 genes en experimentos de sobreexpresión de U1 y 9.166 genes en AMO de U1. Usando Cufflinks, cortes de ajuste de cambio en la expresión de 1,5 veces y valor p <0,01 (Trapnell et al., 2010, Nature Biotechnology 28(5):511-5), 13 genes estaban regulados al alza en AMO de U1, así como 382 y 206 genes en la sobreexpresión de U1 (1,0 y 1,5 µg, respectivamente), mientras 241, 301 y 556 genes, respectivamente, estaban regulados a la baja (Tabla 4). Estos valores indican que los niveles de U1 afectan la estabilidad de los transcritos o la eficiencia de la transcripción de muchos genes. Los cambios en el corte y empalme alternativo se analizaron por MISO, filtros de ajuste de $|\Delta\Psi| \geq 0,3$, factor de Bayes ≥ 10 (Katz et al., 2010, Nat Methods 7(12):1009-15). Esto detectó ~ 5-250 de cada uno de los varios tipos de eventos de corte y empalme alternativo, incluyendo exones omitidos, exones mutuamente exclusivos, intrones retenidos, sitios de corte y empalme en 5' o 3' alternativos, y el primer o último exones alternativos (Tabla 5).

Tabla 3: estadística de RNA-seq

Estadística de mapeo	AMO ctrl	AMO de U1	Ctrl (vector vacío)	Sobreexpresión de U1 (1 µg)	Sobreexpresión de U1 (1,5 µg)
Lecturas totales	97.702.429	94.980.514	226.028.904	185.947.498	207.313.315
Lecturas mapeadas	51.137.948	49.896.982	136.133.739	92.804.677	123.786.737
% de mapeo	52,34 %	52,53 %	60,23 %	49,91 %	59,71 %

Un resumen del número total de lecturas, lecturas mapeadas y el porcentaje de lecturas mapeadas en comparación con el total para cada muestra de RNA-seq.

5 Tabla 4. Análisis de la expresión diferencial

Expresión diferencial ($\pm 1,5$ veces de cambio, $P < 0,01$)	AMO de U1	Sobreexpresión de U1 (1 µg)	Sobreexpresión de U1 (1,5 µg)
Transcritos regulados al alza	13	382	206
Transcritos regulados a la baja	241	301	556

La tabla reporta el número de transcritos regulados al alza y regulados a la baja después de tratamiento con AMO de U1 y sobreexpresión de U1. El análisis se realizó usando el software de expresión diferencial Cufflink (veces de cambio $>1,5$ o $<-1,5$ y valor $p < 0,01$).

10 Tabla 5: análisis MISO de cambios en el corte y empalme alternativo

Eventos de corte y empalme ($ \Delta\Psi \geq 0,3$)	AMO de U1	Sobreexpresión de U1 (1 µg)	Sobreexpresión de U1 (1,5 µg)
Omisión/inclusión de exones	38	130	208
Sitio de corte y empalme alternativo en 5'	5	28	98
Sitio de corte y empalme alternativo en 3'	5	41	40
Exón mutuamente exclusivo	1	7	14
Retención de intrones	12	100	131
Primer exón alternativo	45	144	230
Último exón alternativo	62	143	263

Se reportan las distintas clasificaciones de eventos de corte y empalme usando el software MISO y aplicando un $|\Delta\Psi| \geq 0,3$, factor de Bayes ≥ 10 , y lecturas de exclusión e inclusión ≥ 15 como filtros.

15 Aunque el ARN para secuenciación se seleccionó por oligo(dT), y así se espera que las muestras sobrerrepresenten enormemente los ARNm con colas de poli(A), la baja abundancia de lecturas que contienen poli(A) en 3' no codificada genómicamente de RNA-seq limita la capacidad de identificar los extremos 3'. Como una estrategia más general para identificar los cambios en la longitud, se desarrolló UTRend, una estructura bioinformática que usa un modelo de regresión local para detectar picos de lecturas de RNA-seq. La relación relativa de la intensidad de pico proximal a distal se usó para analizar el patrón del cambio en la longitud del extremo 3' para isoformas de ARNm producidas a partir de cualquier gen dado. Un punto de corte astringente de relación distal a proximal >3 y un valor $p < 0,01$ indica el acortamiento de 3'UTR en la muestra experimental en comparación con el control, mientras aquellas en las que la relación proximal a distal >3 corresponden a alargamiento de 3'UTR. UTRend reveló que la sobreexpresión de U1 incrementó la longitud de las 3'UTR en 1.720 y 1.500 genes para 1,0 y 1,5 µg, respectivamente. Una gran proporción de estos, 1.094, se superpuso a ambos niveles de sobreexpresión de U1, una fuerte validación de los datos de RNA-seq de experimentos biológicos separados. Solo un número relativamente pequeño de genes (82 y 282, a 1,0 y 1,5 µg, respectivamente) tenía acortamiento en las 3'UTR con sobreexpresión de U1. Por el contrario, el AMO de U1 causó un acortamiento de las 3'UTR en 718 genes, mientras solo 28 genes tenían 3'UTR más largas (Figura 23A). La gran mayoría de los genes acortados en AMO de U1 fueron diferentes de aquellos que eran más largos en la sobreexpresión de U1 (superposición de 78 genes). Aunque UTRend subestima el número de genes afectados porque no incluye las UTR con picos únicos, está, no obstante, claro que el alcance de los cambios en la longitud de las 3'UTR es importante. Generalmente, los genes acortados se transcribieron a alto nivel y tenían 3'UTR más largas que el promedio (Figura 23B-23C). Además, no hubo correlación entre la posición donde ocurrió el acortamiento y la distancia desde el sitio de inicio de la transcripción, indicando que factores hasta ahora desconocidos diferencian las 3'UTR del resto del pre-ARNm, haciéndolas más susceptibles a escisión y poliadenilación durante la escasez de *telescripting*.

A partir del alargamiento incitado por la sobreexpresión de U1, se infiere que, respecto a su longitud potencial, es decir, si HeLa fueran quiescentes, 2.126 genes combinados de las dos dosis de sobreexpresión de U1 en estas células cancerosas ya tienen los ARNm con 3'UTR acortadas. A la inversa, un mayor déficit de *telescripting* incitado por AMO de U1, acorta las 3'UTR de 718 genes adicionales. Notablemente, hubo una superposición insignificante entre los genes en los que ocurrieron los cambios en el corte y empalme y aquellos en los que se vio afectada la longitud de las 3'UTR, proporcionando una evidencia adicional de que el corte y empalme y el *telescripting* son eventos dependientes de U1 separados. Aunque el número de cada tipo de cambio del corte y empalme alternativo, efectos en el nivel de los transcritos, y cambios en la longitud de las 3'UTR se evaluaron por diferentes métodos y parámetros y no pueden compararse directamente, los cambios en la longitud de las 3'UTR constituyeron el único tipo más prevalente de cambio en el ARNm que resulta de la variación en la abundancia de U1.

Los ejemplos de alargamiento y acortamiento de las 3'UTR en el RNA-seq, así como las validaciones representativas por 3'RACE, se presentan en las Figuras. 14 y 15. Si nos centramos en el último exón, los patrones de cobertura de las lecturas para la sobreexpresión de U1 (Figura 14) o AMO de U1 (Figura 15) comparados con sus controles respectivos, resaltan el incremento dramático o disminución dependiente de la dosis de las lecturas en la región en 3' (cajas punteadas), respectivamente. Los cambios en la longitud se reflejan tanto por la aparición (Figura 14) o desaparición (Figura 15) del pico o picos distales y por un desplazamiento en la relación de picos distal a proximal (p. ej., SOX4). Obsérvese que el desplazamiento a 3'UTR más largas fue más pronunciado frecuentemente con cantidades crecientes de sobreexpresión de U1. De forma notable, en varios transcritos para la sobreexpresión de U1, las lecturas se extendieron más allá del extremo 3' canónico descrito previamente de los genes (p. ej., BCL2L2), lo que sugiere que el *telescripting* juega un papel importante en la determinación de la terminación de la transcripción. También se indican en las Figuras. 14 y 15 los sitios predichos de TargetScan para miARN expresados en HeLa, ya que la pérdida o restauración de la regulación de miARN es una de las consecuencias importantes de los cambios en la longitud de las UTR.

A continuación, se examinó qué efecto tiene el incremento del *telescripting* en características clave del comportamiento de las células cancerosas, incluyendo la tasa de crecimiento celular, motilidad y capacidad invasiva, usando ensayos *in vitro* cuantitativos establecidos. La sobreexpresión de U1 en células Hela no tuvo efecto en la proliferación o en la viabilidad celular durante hasta 72 hrs en comparación con un vector vacío control (Figura 16A). Sorprendentemente, sin embargo, el incremento de U1 disminuyó significativamente y de forma dependiente de la dosis la migración celular (25-50 %), medida después de 24 hrs por un ensayo estándar usado para determinar el número de células que atraviesan una membrana de policarbonato porosa en respuesta a un quimioatrayente (mayor concentración de suero) (Kramer et al., 2012, Mutat Res, 752(1):10-24; Tsujii et al., 1998, Cell 93(5):705-16). Otra medida clave de oncogenicidad es la capacidad de las células tumorales de metastatizar, que, además de la migración, requiere la capacidad de atravesar una capa de matriz extracelular de membrana basal (p. ej., Matrigel) (Albini et al., 1987, Cancer Res 47(12):3239-45; Albini y Noonan, 2010, Curr Opin Cell Biol 22(5):677-89). La sobreexpresión de U1 también disminuyó fuertemente la capacidad invasiva un 25-65 % de una manera dependiente de la dosis (Figuras 16B-16E).

También se examinó el efecto de una disminución moderada del nivel de U1 en los mismos ensayos. En el rango de dosis baja de AMO de U1 (0,05-0,25 nmoles), no hubo un efecto en la proliferación de las células HeLa después de 24 hrs (Figura 17). Se observó un incremento moderado (~ 35 %) en el número de células en comparación con el AMO control después de 2-3 días a 0,25 nmoles de AMO de U1. El crecimiento celular se inhibió a la dosis más alta de AMO de U1 (1,0 nmol), cuando fue evidente el acortamiento más drástico de ARNm (p. ej., en intrones tempranos) y alguna inhibición del corte y empalme (Berg et al., 2012, Cell 150(1):53-64). De forma notable, los niveles bajos de AMO de U1 (0,05-0,25 nmoles) causaron un aumento del 400-500 % en la migración celular en 24 hrs (Figuras 17B y 17D). Como la dosis idéntica de AMO de U1 no tuvo efecto en el número de células en el mismo punto de tiempo (Figura 17A), esto refleja un incremento verdadero en la actividad de migración. También se observó un incremento dramático en la capacidad invasiva (400-500 %), similar en dependencia de la dosis al observado para la migración, con una actividad máxima a 0,25 nmoles de AMO de U1 (Figuras 17C y 17E). Como un control, se realizaron los mismos experimentos con AMO de U2, que interfiere con esta función de la RNPSn en el corte y empalme. No se observó ningún aumento en ninguna de las actividades medidas, indicando que los efectos en la oncogenicidad son específicos de la función de *telescripting* de U1 (Figura 24).

También se observaron efectos similares en otras células cancerosas ensayadas, incluyendo adenocarcinoma de pulmón humano (A549) y adenocarcinomas de mama (MCF-7 y MB-231). El AMO de U1 aumentó la migración de A549, MCF-7 y MB-231 un 58-72 % e incrementó la invasión de A549 y MB-231 un 53-64 %. Aunque no se pretende la vinculación a ninguna teoría particular, la menor magnitud de los efectos en comparación con HeLa podría reflejar diferencias específicas de célula, incluyendo la eficiencia de la transfección, pero también se observa que estas líneas celulares cancerosas mostraron una agresividad mayor en la línea base. Con el fin de discernir de forma fiable los cambios posteriores a la transfección, se sembraron en placas 6 veces menos células en comparación con HeLa, disminuyendo de esta manera la probabilidad de observar incrementos adicionales en la agresividad. De manera similar a HeLa, la sobreexpresión de U1 atenuó la migración e invasión de estas líneas celulares ~ 50 % en

comparación con los niveles control (Tabla 6). Por lo tanto, el efecto de los cambios en U1 en la oncogenicidad de las células cancerosas es un fenómeno general.

Tabla 6: mediciones de migración e invasión en líneas celulares cancerosas adicionales

Líneas celulares	AMO de U1 (0,25 nmoles)					Sobreexpresión de U1						
	Índice de migración (% del ctl)	Índice de invasión (% del ctl)	Proliferación (% de incremento)			Índice de migración (% del ctl)		Índice de invasión (% del ctl)		Proliferación (% de incremento)		
			24 hr	48 hr	72 hr	1 µg	1,5 µg	1 µg	1,5 µg	24 hr	48 hr	72 hr
HeLa (cáncer de cuello uterino)	500	463	0	38	38	74	52	67	37	Sin efecto		
A-549 (adenocarcinoma pulmonar)	172	164	0	13	15	66	47	92	91	Sin efecto		
MCF-7 (adenocarcinoma de mama)	158	n/a	0	10	12	n/a						
MB-231 (adenocarcinoma de mama)	170	153	0	5	7	57	54	63	72	Sin efecto		

- 5 Se ensayó el efecto de los cambios en los niveles de U1 en el comportamiento del cáncer en varias líneas celulares adicionales. La tabla resume la proliferación, índice de migración e invasión para las células tratadas en comparación con el control después de AMO de U1 o sobreexpresión de U1 (ambos 1 y 1,5 µg).

En la búsqueda de la explicación potencial para los cambios en la agresividad del cáncer, se analizaron los genes en los que se detectaron los cambios en la longitud de las 3'UTR. Como se esperaba para un número tan grande de genes, el análisis de la ontología génica y de rutas (DAVID) (Huang 2009, Nat Protoc 4(1):44-57) mostró que el acortamiento de las 3'UTR afectó a genes con funciones ampliamente diversas. No obstante, los genes implicados en las rutas del cáncer estaban significativamente enriquecidos (Tabla 7). Así, se buscó en dos bases de datos de genes de cáncer: Sanger (Futreal et al., 2004, Nat Rev Cancer 4(3):177-83) y UCSF, que incluyen protooncogenes cuya regulación al alza por varios mecanismos, incluyendo mutación, traslocación cromosómica, o pérdida de represión de miARN, está ligada al cáncer. Esto reveló que 142 de 456 genes del cáncer en estas bases de datos que se expresan en HeLa tenían sus 3'UTR afectadas por los cambios en el nivel de U1. La Figura 21 lista estos genes según el nivel de U1 que causó el acortamiento de sus 3'UTR (AMO de U1; 45 genes) o revirtió sus ya acortadas 3'UTR a formas más largas (85 y 78 para 1 y 1,5 µg, respectivamente). Estos datos muestran que cada disminución incremental en el nivel de U1 causa el acortamiento de las 3'UTR en un número creciente de oncogenes y un fenotipo de cáncer correspondientemente más agresivo. Los genes afectados listados en la Figura 21 y la Tabla 7 tienen funciones conocidas en el inicio y la progresión del cáncer, incluyendo regulación del ciclo celular (CDC25A, SKP1), apoptosis (BAG1, BCL2), adhesión celular (CD164, TJP1), motilidad (ACTN1, FAP), remodelado de la matriz extracelular (TIMP2, TIMP3), señalización (WNTSA, EGFR), transcripción (RARA, NFKB1), metástasis (EWSR1, APC, MDM2), y progresión tumoral (SLCA16A1, CAPZA2).

Tabla 7: análisis de la ruta por DAVID

Término	Genes afectados	Veces de enriquecimiento	Benjamini
Rutas en cáncer	WNT5A, FGFR1, PTGS2, ARNT2, TFG, NFKB1, TCF7L2, CDC42, FOS, CUL2, BCL2, SLC2A1, CASP8, RALB, RHOA, RARA, FGF2, TRAF4, APC, COL4A4, EGFR, DVL3, IL6, CTBP2, MAP2K1, BRAF, TGFB2, FZD1, SKP2, BRCA2, FZD5, BIRC3, COL4A6, CDK2, CCDC6, GSK3B, VEGFA, MDM2, LAMC1, IKBKB	13,9	$3,57 \times 10^{-30}$
Ruta de señalización MAPK	FGFR1, GNA12, PPP3R1, NFKB1, SRF, ATF2, PATHWAY HSPA1L, CDC42, FOS, DUSP14, MAP3K2, PPP3CA, CHP, RAPGEF2, FGF2, EGFR, MAP2K1, BRAF, TAOK1, TAOK3, TGFB2, TAB2, DUSP4, MAP4K4, DUSP3, RPS6KA3, ARRB2, DUSP1, IKBKB, GADD45A, DUSP6	14,6	$7,27 \times 10^{-24}$
Ruta de señalización WNT	WNT5A, DVL3, CTBP2, VANGL1, PPP2R5C, FZD1, PPP3R1, FZD5, TCF7L2, SENP2, CSNK2A1, GSK3B, PPP2CA, NFAT5, RHOA, LRP6, CAMK2D, PPP2R5E, S1AH1, PPP3CA, CHP, TBL1X, APC	19,1	$3,39 \times 10^{-20}$
Regulación del citoesqueleto de actina	EGFR, GNA13, FGFR1, MAP2K1, BRAF, MYLK3, WASF1, GNA12, SSH2, ABI2, ACTN1, MYH9, ARHGEF12, PPP1CB, NCKAP1, CDC42, DOCK1, ITGB8, RHOA, CYFIP1, FGF2, F2R, APC	14,1	$2,94 \times 10^{-17}$
Unión adherente	EGFR, PTPRJ, FGFR1, WASF1, TGFB2, ACTN1, CTNND1, FER, ACP1, TCF7L2, CDC42, TJP1, CSNK2A1, FYN, RHOA, PTPN1, YES1, INSR	23,2	$3,66 \times 10^{-17}$
Adhesión focal	EGFR, COL4A4, CAV1, BRAF, MAP2K1, MYLK3, ACTN1, BIRC3, PPP1CB, COL4A6, CDC42, PDPK1, DOCK1, FYN, ITGB8, GSK3B, BCL2, VEGFA, RHOA, LAMC1	13	$2,22 \times 10^{-14}$
Ciclo celular	ANAPC13, YWHAB, SKP2, PRKDC, CDK7, MCM4, SMC3, CDK2, CDC25A, RAD21, YWHAH, GSK3B, YWHAQ, MDM2, SMC1A, GADD45A, STAG2	12,2	$6,55 \times 10^{-12}$
Uniones firmes	F11R, PARD6B, MAGI3, VAPA, CNKSR3, PRKCI, ACTN1, MYH9, CDC42, TJP1, CSNK2A1, PPP2CA, RHOA, YES1, PPP2R2A	14,1	$2,49 \times 10^{-11}$
Apoptosis	IRAK1, PPP3R1, NFKB1, BIRC3, CAPN1, CASP7, BCL2, RIPK1, PRKAR1A, CASP8, CHP, PPP3CA, IKBKB	15,9	$4,43 \times 10^{-10}$
Ruta de señalización de p53	EI24, SERPINB5, SHISA5, CASP8, MDM2, S1AH1, RCHY1, PMAIP1, PERP, SESN2, GADD45A, CDK2	16,5	$1,01 \times 10^{-9}$

En la tabla se muestran los términos relacionados con el cáncer identificados por el análisis DAVID de rutas de KEGG enriquecidas. Se muestran los genes afectados para cada categoría, así como las veces de enriquecimiento. Los valores p se corrigieron para ensayo múltiple usando el método de Benjamini-Hochberg.

Se han descrito numerosos ejemplos de cánceres debidos a la regulación al alza de oncogenes que resulta de la pérdida de la represión de miARN, bien porque el miARN relevante está regulado a la baja o se ha eliminado su diana en la 3'UTR (Ebert y Sharp, 2012, Cell 149(3):515-24; Mendell y Olson, 2012, Cell 148(6):1172-87; Nana-Sinkam y Croce, 2011, Mol Oncol 5(6):483-91). De forma importante, los cambios en el nivel de U1 recapitularon la misma eliminación o restauración de la diana 3'UTR del miARN en muchos genes (Figuras. 14, 15 y 21). Por ejemplo, el acortamiento en las 3'UTR de CCND1 (ciclina D1) y RAB10 (miembro de la familia de oncogenes RAS) se encuentra

en numerosas líneas celulares cancerosas, que entre muchas otras, elimina sitios diana para miRs-15/16 y miRs-103/107, respectivamente, lo que da lugar a un incremento dramático en el nivel de proteínas de estos oncogenes (Lin et al., 2012, Nucleic Acids Res 40(17):8460-71; Mayr y Bartel, 2009, Cell 138(4):673-84). Como se muestra en la Figura 14, la 3'UTR de RAB10 ya está acortada en las células HeLa, y la sobreexpresión de U1 revierte el acortamiento de la 3'UTR y restaura los sitios de unión correspondientes de miARN. Para CDC25A, una fosfatasa esencial para la transición G1-S, el acortamiento de la 3'UTR elimina sitios de unión de miARN, incluyendo let-7, miR-15 y miR-21 (Figura 15). Un incremento en la proteína CDC25A, debido al alivio de la represión mediada por miARN, exacerba la formación de quistes hepáticos y el cáncer de colon (Johnson et al., 2007, Cancer Res 67(16):7713-22; Wang et al., 2009, Cancer Res 69(20):8157-65; Lee et al., 2008, J Clin Invest 118(11):3714-24). El receptor del factor de crecimiento epidérmico (EGFR) está sobreexpresado en una variedad de tumores (Arteaga, 2001, J Clin Oncol 19(18 Supl):32S-40S) y AMO de U1 elimina la secuencia diana de miR-7 en la 3'UTR (Figura 15), lo que causa una capacidad invasiva y metástasis incrementadas (Webster et al., 2009, J Biol Chem 284(9):5731-41). Se encontró que el *telescoping* también regula las 3'UTR de tirosina quinasas Src no de receptor (YES1, FYN, y LYN), que son dianas aguas abajo de EGFR implicadas en la señalización para el crecimiento celular y formación de unión firme, y son altamente oncogénicas cuando se disrumpe su regulación por miARN (Figura 21) (Majid et al., 2011, Cancer Res 71(7):2611-21; Sommer et al., 2005, J Steroid Biochem Mol Biol 97(3):219-29). De forma similar, un polimorfismo en la 3'UTR de KRAS incluido en la misma región que se elimina por una disminución de U1, altera la unión de let-7 y es pronóstico para la agresividad del cáncer de mama (Paranjape et al., 2011, Lancet Oncol 12(4):377-86) (Figura 14). También se ha mostrado que la regulación a la baja de los miARN causa cáncer debido a la pérdida de represión del ARNm específico que codifica oncogenes (Nicoloso et al., 2009, Nat Rev Cancer 9(4):293-302). Por ejemplo, la migración y la invasión de las células de cáncer de mama están aumentadas por la regulación a la baja de miR-339-5p (Wu et al., 2010, BMC Cancer 10:542), para el que se pierden los sitios diana en BCL6 cuando se reduce U1 (Figura 15). Asimismo, la regulación a la baja de miR-335 regula al alza SOX4, que promueve la transición epitelial a mesenquimal y aumenta las metástasis del cáncer de mama (Tavazoie et al., 2008, Nature 451(7175):147-52). La 3'UTR de SOX4 en HeLa ya está acortada, pero el sitio de unión de miR-335 se restaura por la sobreexpresión de U1 (Figura 14).

Consistente con la expectación de que el acortamiento de 3'UTR regularía al alza generalmente la producción de proteínas a partir de los ARNm correspondientes, que son numerosos, AMO de U1 bajo incrementó el rendimiento de la traducción un ~ 20 % según se mide por incorporación de [³⁵S]metionina durante 1 hr. Por el contrario, la sobreexpresión de U1 disminuyó la síntesis de proteínas un 30 % (Figura 18). Así, hay un gran diferencial, ~ 50 %, en la actividad de síntesis de proteínas entre las células cancerosas más agresivas (AMO de U1) y las más atenuadas (sobreexpresión de U1, 1.5 µg). Como controles, la cicloheximida inhibió claramente la síntesis de proteínas, y AMO de U2, que no tenía efecto en el fenotipo del cáncer, no afectó la traducción (Figura 18). Notablemente, la sobreexpresión de un factor de inicio de la traducción es oncogénica (Furic et al., 2010, Proc Natl Acad Sci USA 7(32):14134-9), y una inhibición moderada de la síntesis de proteínas del ~ 50 % es suficiente para inhibir el crecimiento de las células cancerosas (Medrano y Pardee, 1980, Proc Natl Acad Sci USA 77(7):4123-6).

Además de muchos oncogenes, los niveles de U1 causaron cambios en la longitud de 3'UTR en numerosos otros genes que podrían contribuir al fenotipo del cáncer. Se observó un número mucho más pequeño de eventos de corte y empalme alternativo, así como cambios en el nivel de ARNm en muchos genes y algunos de estos también juegan probablemente un papel en el fenotipo relacionado con el cáncer. Los eventos de corte y empalme alternativo podrían ser una consecuencia directa del cambio en el nivel de U1, debido a su papel en el corte y empalme, pero, sin pretender la vinculación a ninguna teoría particular, también podrían ser secundarios a los cambios en la longitud de 3'UTR en muchos otros reguladores del corte y empalme. Sin pretender la vinculación a ninguna teoría particular, los cambios en el nivel de ARNm podrían ser el resultado de cambios en la transcripción, ya que muchos de los genes afectados son reguladores de la transcripción. Sin embargo, también podrían ser un resultado directo de cambio en *telescoping*, ya que podría afectar fuertemente la estabilidad del ARNm (p. ej., por la eliminación de un elemento estabilizador).

Los experimentos descritos en un sistema de células cancerosas definido y los ensayos cuantitativos revelaron que cambios relativamente pequeños en *telescoping* podrían modular profundamente el fenotipo del cáncer asociado con la agresividad tumoral. Varias líneas de evidencia apoyan la relevancia fisiológica de las informaciones obtenidas a partir de los experimentos de modulación de U1 y sugieren fuertemente que los cambios fenotípicos están mediados principalmente por la regulación del *telescoping*. En primer lugar, el déficit de *telescoping* recapituló el acortamiento general de 3'UTR en líneas de células cancerosas (Mayr y Bartel, 2009, Cell 138(4):673-84; Lin et al., 2012, Nucleic Acids Res 40(17):8460-71; Elkon et al., 2012, Genome Biol 13(7):R59; Fu et al., 2011, Genome Res 21(5):741-7) y en tumores (Lembo et al., 2012, PLoS One 7(2):e31129), y produjo la misma pérdida de represión de miARN de oncogenes que subyace a muchos cánceres. En segundo lugar, los cambios en el fenotipo correspondieron tanto a la direccionalidad como a la magnitud de los cambios en la longitud de las 3'UTR, incluyendo el número de oncogenes y el grado del cambio de las 3'UTR (p. ej., desplazamiento dependiente de la dosis a 3'UTR más largas; (Figuras 14, 15 y 21). En tercer lugar, a diferencia de AMO de U1, AMO de U2, que también afecta el corte y empalme y el nivel de ARNm, no afectó el fenotipo del cáncer (Figura 24), lo que sugiere fuertemente que el fenotipo del cáncer está

causado por la actividad única de U1 en *telescripting*. Apoyando adicionalmente la relevancia de las presentes observaciones para la oncología clínica, también se encontró que >45 genes para los que se observó acortamiento de 3'UTR estaban acortados en cánceres de pulmón y mama lo que se correlacionó con mal pronóstico (Lembo et al., 2012, PLoS One 7(2):e31129).

5 Se desconoce lo que causa el déficit de *telescripting* en el cáncer, pero no depende necesariamente de la disminución de U1 en términos absolutos y también podría estar causado por una regulación al alza de la transcripción, así como por un incremento en factores de escisión/poliadenilación. De hecho, sin pretender la vinculación a ninguna teoría particular, pequeños cambios en la transcripción, que crean una escasez de U1 respecto a los transcritos nacientes normalmente protegidos que necesita proteger sin disminuir necesariamente los niveles de U1 (Berg et al., 2012, Cell 10 150(1):53-64), podrían ser suficientes para modular el *telescripting* y causar cambios en la longitud del ARNm y regulación de las isoformas (Berg et al., 2012, Cell 150(1):53-64). Se han descrito varios factores, además de U1, que pueden influir en la longitud de las 3'UTR, incluyendo componentes de la maquinaria de escisión y poliadenilación [Cstf64, CFI(m)68, PABPN1] (Di Giammartino et al., 2011, Mol Cell 43(6):853-66; de Klerk et al., 2012, Nucleic Acids Res 40(18):9089-101; Yao et al., 2012, Proc Natl Acad Sci USA 109(46):18773-8; Takagaki et al., 1996, Cell 87(5):941- 15 52; Martin et al., 2012, Cell Rep 1(6):753-63). Dado su papel central en la expresión génica, es probable que el *telescripting* sea un proceso altamente regulado, y que la longitud de los ARNm se determinará por el equilibrio entre la maquinaria de escisión y poliadenilación proacortamiento y la capacidad de *telescripting*, que es altamente dependiente del rendimiento transcripcional respecto a U1 disponible y probablemente varía en diferentes tipos de células y estados fisiológicos. En la presente memoria se propone que el déficit de *telescripting* también explica el acortamiento de 3'UTR que se observa en las células proliferantes, así como en las células inmunes y neuronas activadas. Es común a todas estas un incremento en la transcripción y en el uso de APA proximales (Ji et al., 2011, Mol Syst Biol 7:534).

Mediante la determinación de la longitud de ARNm de miles de genes, el *telescripting* es un regulador principal de APA y de la expresión génica, controlando la disponibilidad de dianas de los miARN y otros elementos reguladores en las 3'UTR y cambio de isoformas que codifican proteínas que resultan de PCPA en intrones. Una pérdida de dianas de miARN en una escala tal también modularía probablemente la robustez de la regulación del miARN en las dianas remanentes en otros ARNm (Ebert y Sharp, 2012, Cell 149(3):515-24) y reverberaría a través de la red de la miríada de factores que regulan la función del ARNm. Adicionalmente, las 3'UTR también están altamente enriquecidas en elementos reguladores, tales como restos ricos en AU y sitios para proteínas de unión a ARN que impactan en cada 25 30 aspecto de la fisiología de los ARNm (Chen y Shyu, 1995, Trends Biochem Sci 20(11):465-70), y estos también se verían afectados por los cambios en la longitud de las 3'UTR.

A partir de los estudios presentados en la presente memoria surge una nueva perspectiva sobre la implicación de los oncogenes en el cáncer. Aunque la transformación celular puede iniciarse por una mutación somática en un único oncogén, los descubrimientos presentados en la presente memoria indican que una célula cancerosa establecida 35 también tiene un gran número de oncogenes que están desregulados cotranscripcionalmente por déficit de *telescripting* y sería indetectable por secuenciación genómica. De forma importante, los datos presentados en la presente memoria muestran que los cambios en el nivel de U1 solo pueden modificar fuertemente la longitud de los ARNm y su fenotipo del cáncer asociado. Por lo tanto, en la presente memoria se demuestra que la modulación del *telescripting* tiene potencial como un tratamiento de amplio espectro para el cáncer y otros fenotipos asociados con el 40 acortamiento de los ARNm.

LISTA DE SECUENCIAS

5 <110> The Trustees of the university of Pennsylvania
 Dreyfuss, Gideon
 Wan, Lili
 Oh, Jung Min
 Pinto, Anna Maria
 Berg, Michael George

10 <120> La RNPsn U1 regula la expresión génica y modula la oncogenicidad
 <130> 046483-6076-00-WO.602632

15 <150> 61/668,168
 <151> 05-07-2012
 <160> 63

20 <170> PatentIn versión 3.5
 <210> 1
 <211> 9
 <212> ADN
 <213> Secuencia Artificial

25 <220>
 <223> secuencia de reconocimiento del sitio de corte y empalme en 5' de tipo salvaje de U1

30 <400> 1
 acttacctg 9

35 <210> 2
 <211> 9
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> U1-A: secuencia de reconocimiento del sitio de corte y empalme en 5' mutada de U1

40 <400> 2
 ctgctgttg 9

45 <210> 3
 <211> 9
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> U1-B: secuencia de reconocimiento del sitio de corte y empalme en 5' mutada de U1

<400> 3
 tgaatggac 9

55 <210> 4
 <211> 9
 <212> ADN
 <213> Secuencia Artificial

60 <220>
 <223> U1-C: secuencia de reconocimiento del sitio de corte y empalme en 5' mutada de U1

<400> 4
 ctacagatg 9

<210> 5
 <211> 9
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> U1-D: secuencia de reconocimiento del sitio de corte y empalme en 5' mutada de U1
 <400> 5
 10 caccctac 9
 <210> 6
 <211> 25
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 20 <400> 6
 cctcttacct cagttacaat ttata 25
 <210> 7
 <211> 25
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 30 <400> 7
 ggtatctccc ctgccaggta agtat 25
 <210> 8
 <211> 25
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 40 <223> Sintetizado químicamente
 <400> 8
 tgataagaac agatactaca cttga 25
 45 <210> 9
 <211> 25
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> Sintetizado químicamente
 <400> 9
 55 aaccttctct ctttcatac aacac 25
 <210> 10
 <211> 24
 <212> ADN
 <213> Secuencia Artificial
 60 <220>
 <223> Sintetizado químicamente
 <400> 10
 65 tcgtatttt ccttactcat aagt 24

<210> 11
 <211> 25
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> Sintetizado químicamente
 <400> 11
 10 gttaacctct acgccaggta agtat 25
 <210> 12
 <211> 25
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 20 <400> 12
 gataagaaca gatactacac ttgta 25
 <210> 13
 <211> 25
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 30 <400> 13
 aacttgctct ccttcaaac aacac 25
 <210> 14
 <211> 25
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 40 <223> Sintetizado químicamente
 <400> 14
 gattggtttt ccttactcat tagtt 25
 45 <210> 15
 <211> 25
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> Sintetizado químicamente
 <400> 15
 55 ggatattttc ttctaagaaag gaatc 25
 <210> 16
 <211> 25
 <212> ADN
 <213> Secuencia Artificial
 60 <220>
 <223> Sintetizado químicamente
 <400> 16
 65 aataccacag ttattgaag actac 25

<210> 17
 <211> 59
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> Sintetizado químicamente
 <400> 17
 10 cgtatcgctt cctcgcgcc atcagacgag tgcgtgctcg agcggccgcc cgggcaggt 59
 <210> 18
 <211> 58
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 20 <400> 18
 ctatgcgctt tgccagcccg ctcagacgag tgcgtggcag cgtggtcgcg gccgaggt 58
 <210> 19
 <211> 59
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 30 <400> 19
 cgtatcgctt cctcgcgcc atcagacgct cgacagctcg agcggccgcc cgggcaggt 59
 <210> 20
 <211> 58
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 40 <223> Sintetizado químicamente
 <400> 20
 ctatgcgctt tgccagcccg ctcagacgct cgacaggcag cgtggtcgcg gccgaggt 58
 <210> 21
 <211> 20
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> Sintetizado químicamente
 <400> 21
 55 cagtctgcca ttgtgacgtg 20
 <210> 22
 <211> 20
 <212> ADN
 <213> Secuencia Artificial
 60 <220>
 <223> Sintetizado químicamente
 <400> 22
 65 ccaatgggag agccctaaac 20
 <210> 23

<211> 19
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> Sintetizado químicamente

<400> 23
 ctgcaatggg agttgggag 19

10 <210> 24
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

15 <220>
 <223> Sintetizado químicamente

20 <400> 24
 gatgcgttg attctgccca c 21

25 <210> 25
 <211> 25
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintetizado químicamente

30 <400> 25
 gaggagctct ccaaactgat cagcg 25

35 <210> 26
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintetizado químicamente

40 <400> 26
 ctgatcagcg gcttgacca 20

45 <210> 27
 <211> 27
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> Sintetizado químicamente

<400> 27
 cgagtatgag tcgactctag actgccg 27

55 <210> 28
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

60 <220>
 <223> Sintetizado químicamente

<400> 28
 ctctagactg ccgcgacgtg g 21

65 <210> 29

<211> 40
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> Sintetizado químicamente
 <400> 29
 ctgatctaga ggtaccggat ccttttttt tttttttt 40
 10 <210> 30
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 15 <220>
 <223> Sintetizado químicamente
 <400> 30
 20 ctgatctaga ggtaccggat cc 22
 <210> 31
 <211> 25
 <212> ADN
 25 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 30 <400> 31
 gctaactcag gggtgcata ggcac 25
 <210> 32
 <211> 25
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente
 40 <400> 32
 cgtccttctc gccctcctc tcctc 25
 <210> 33
 45 <211> 23
 <212> ADN
 <213> Secuencia Artificial
 <220>
 50 <223> Sintetizado químicamente
 <400> 33
 ggttggttt ctttaattgc atc 23
 55 <210> 34
 <211> 22
 <212> ADN
 <213> Secuencia Artificial
 60 <220>
 <223> Sintetizado químicamente
 <400> 34
 gatgattcc attgctgcca cg 22
 65 <210> 35

<211> 25
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> Sintetizado químicamente

<400> 35
 gtcctgaaac cacagcagtg cccag 25

10 <210> 36
 <211> 25
 <212> ADN
 <213> Secuencia Artificial

15 <220>
 <223> Sintetizado químicamente

20 <400> 36
 ctcagcactg acggccttgt tgagt 25

25 <210> 37
 <211> 24
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintetizado químicamente

30 <400> 37
 cagacgatga gagaacaccc gatg 24

35 <210> 38
 <211> 23
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintetizado químicamente

40 <400> 38
 tggcacctct gtgggcctgt ggc 23

45 <210> 39
 <211> 27
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> Sintetizado químicamente

<400> 39
 ctgttgcttc cactgcttca cattggc 27

55 <210> 40
 <211> 22
 <212> ADN
 <213> Secuencia Artificial

60 <220>
 <223> Sintetizado químicamente

<400> 40
 aaggataagc agtgtgaaca ag 22

65 <210> 41

<211> 20
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> Sintetizado químicamente

<400> 41
 tctagatctc ccatcacggc 20

10 <210> 42
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

15 <220>
 <223> Sintetizado químicamente

<400> 42
 20 cagcagtgcc gaaagacata 20

<210> 43
 <211> 26
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente

<400> 43
 30 cgctatcag tctgtccctg gactcc 26

<210> 44
 <211> 25
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente

<400> 44
 40 ccttccttg gtaaggagc tgccc 25

<210> 45
 <211> 22
 45 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Sintetizado químicamente

<400> 45
 50 ccggaattgc tgagtaagct gg 22

<210> 46
 <211> 23
 55 <212> ADN
 <213> Secuencia Artificial

<220>
 60 <223> Sintetizado químicamente

<400> 46
 65 ggaaggcata tactggcaaa tgc 23

5 <210> 47
 <211> 24
 <212> ADN
 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente

 10 <400> 47
 ccttagctc ctcgaattta acag 24

 <210> 48
 <211> 24
 15 <212> ADN
 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente
 20 <400> 48
 ccttgatgaa ctcctcaga ggac 24

 <210> 49
 25 <211> 21
 <212> ADN
 <213> Secuencia Artificial

 <220>
 30 <223> Sintetizado químicamente

 <400> 49
 ggttggtgat ggtggtactg g 21

 35 <210> 50
 <211> 23
 <212> ADN
 <213> Secuencia Artificial

 40 <220>
 <223> Sintetizado químicamente

 <400> 50
 45 gatttcgcct tcacttcct gtc 23

 <210> 51
 <211> 29
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> Sintetizado químicamente

 <400> 51
 55 gtgtggatgt agtacacatg gatgaaaac 29

 <210> 52
 <211> 28
 <212> ADN
 60 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente

 65 <400> 52
 ttattgtaca ccaggacctt ctccacag 28

5 <210> 53
 <211> 28
 <212> ADN
 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente

 10 <400> 53
 caactatggc tgtggagaag gtcctggt 28

 <210> 54
 <211> 31
 15 <212> ADN
 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente

 20 <400> 54
 tctgacctgg agacgaaact gtagagtatc t 31

 <210> 55
 25 <211> 22
 <212> ADN
 <213> Secuencia Artificial

 <220>
 30 <223> Sintetizado químicamente

 <400> 55
 caacagacat tgctgaagaa ac 22

 35 <210> 56
 <211> 18
 <212> ADN
 <213> Secuencia Artificial

 40 <220>
 <223> Sintetizado químicamente

 <400> 56
 45 atttggatg actgcggc 18

 <210> 57
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

 50 <220>
 <223> Sintetizado químicamente

 <400> 57
 55 cttaggagc tgctgctgat a 21

 <210> 58
 <211> 21
 <212> ADN
 60 <213> Secuencia Artificial

 <220>
 <223> Sintetizado químicamente

 65

<400> 58
 gctcctacct cttccaagca g 21

5 <210> 59
 <211> 19
 <212> ADN
 <213> Secuencia Artificial

10 <220>
 <223> Sintetizado químicamente

<400> 59
 acaaaacaca gccccagc 19

15 <210> 60
 <211> 24
 <212> ADN
 <213> Secuencia Artificial

20 <220>
 <223> Sintetizado químicamente

<400> 60
 tgggtgtaat gtcagtgta atgc 24

25 <210> 61
 <211> 22
 <212> ADN
 <213> Secuencia Artificial

30 <220>
 <223> Sintetizado químicamente

<400> 61
 gcaagtgaa ggaagcagca gg 22

35 <210> 62
 <211> 21
 <212> ADN
 <213> Secuencia Artificial

40 <220>
 <223> Sintetizado químicamente

45 <400> 62
 tgatcacgaa ggtggtttc c 21

<210> 63
 <211> 19
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> Sintetizado químicamente

55 <400> 63
 gcacatccgg agtgcaatg 19

REIVINDICACIONES

- 5 1. Un ácido nucleico que codifica un ARNsn de U1 endógeno, para uso en el tratamiento de una enfermedad o trastorno, en donde la enfermedad o trastorno se selecciona del grupo que consiste en cáncer, enfermedad autoinmune, inflamación, neurológica, y cualquier combinación de las mismas.
2. El ácido nucleico que codifica un ARNsn de U1 endógeno para uso en un método de la reivindicación 1, que comprende además un plásmido de expresión de U1.
- 10 3. Una molécula de ácido nucleico que incrementa el nivel de partículas de RNPsn U1 endógena, para uso en el tratamiento de una enfermedad o trastorno, en donde la enfermedad o trastorno se selecciona del grupo que consiste en cáncer, enfermedad autoinmune, inflamación, neurológica, y cualquier combinación de las mismas y en donde la molécula de ácido nucleico es un plásmido que contiene el gen de ARNsn de U1 bajo el control de su promotor nativo.

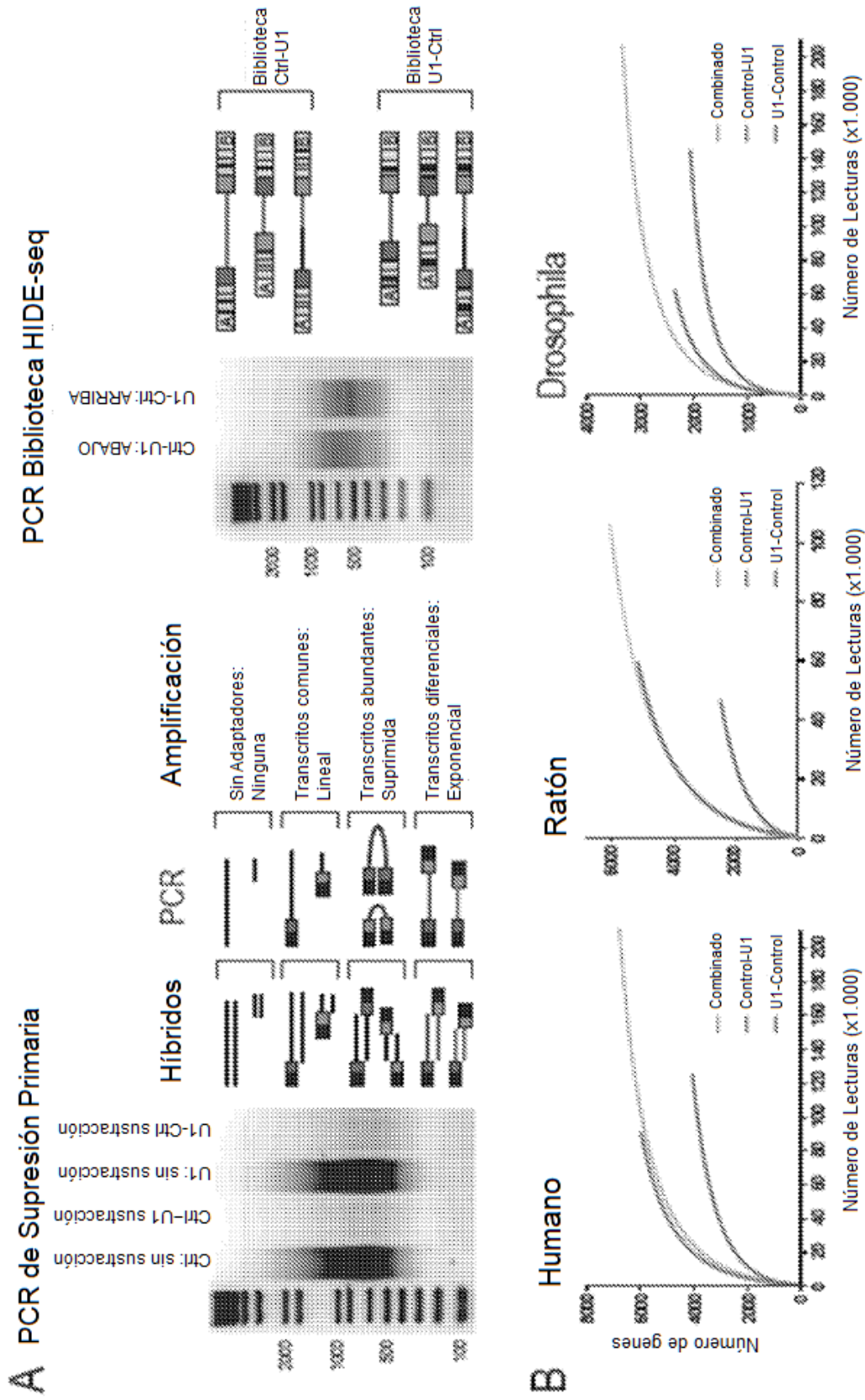


Figura 1

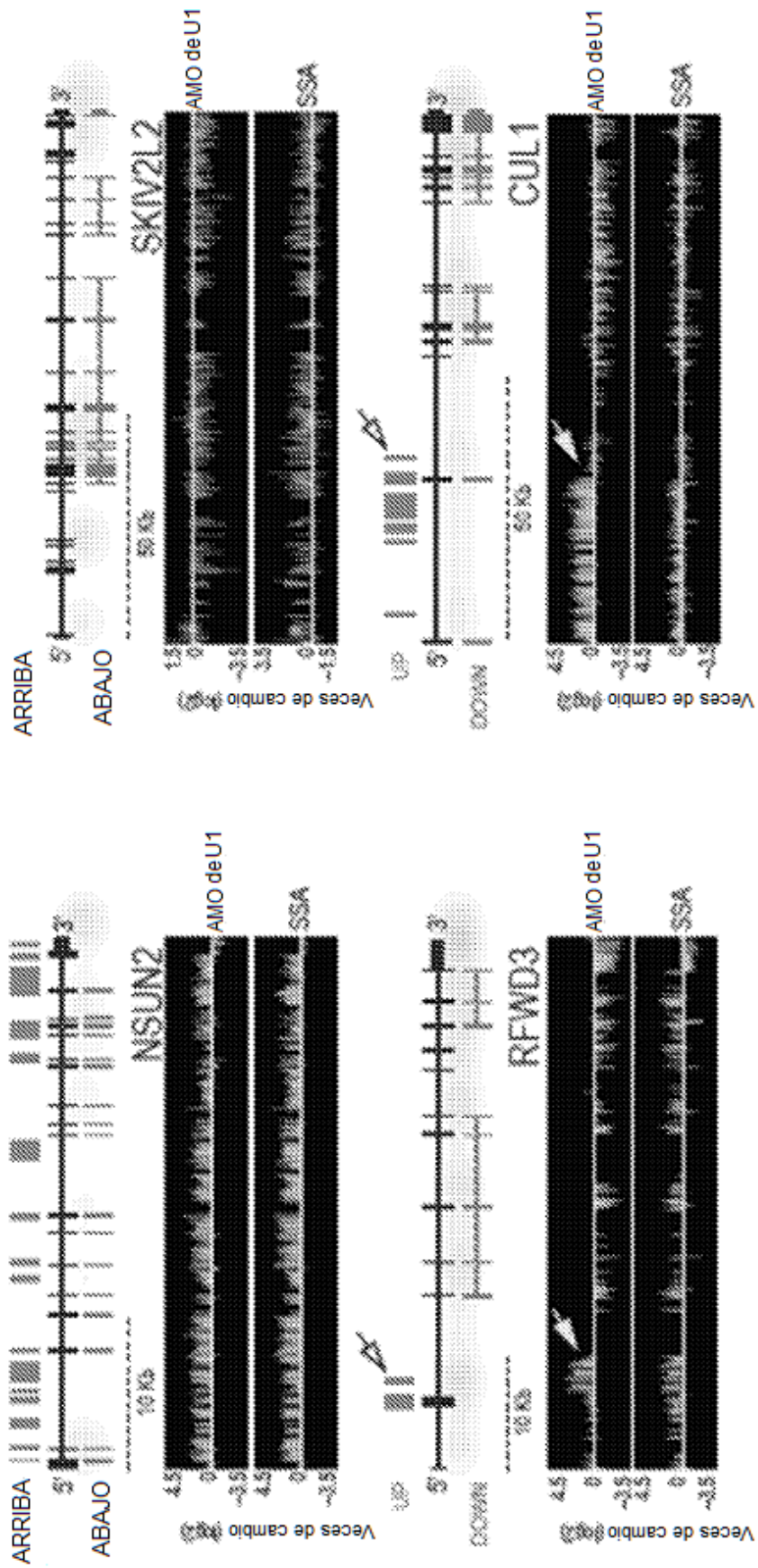
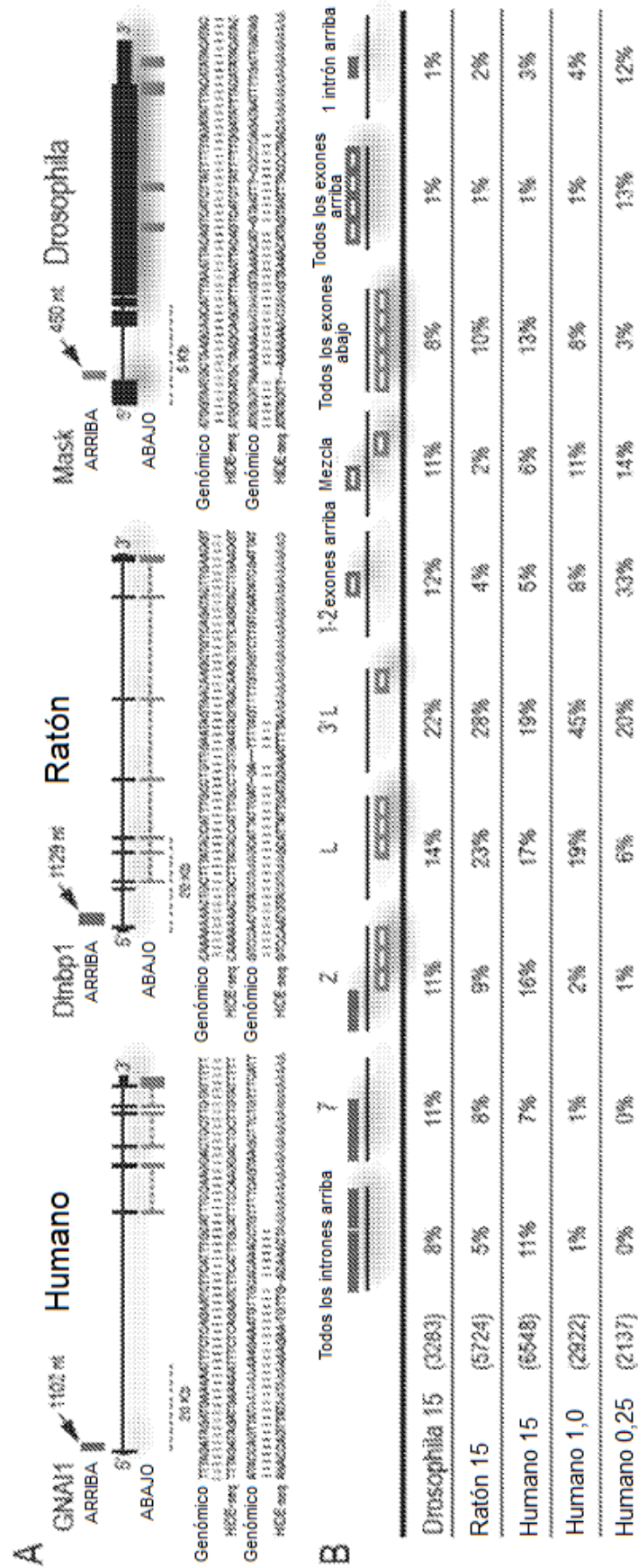


Figura 1C



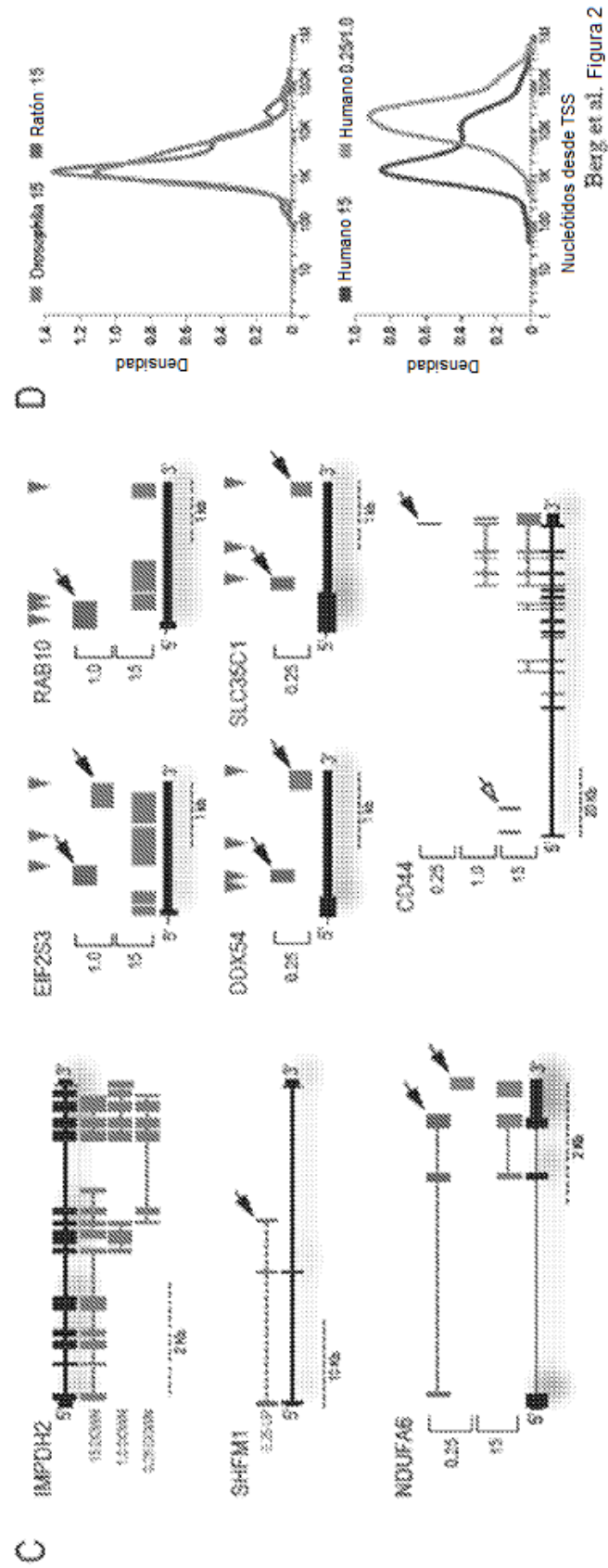


Figura 2

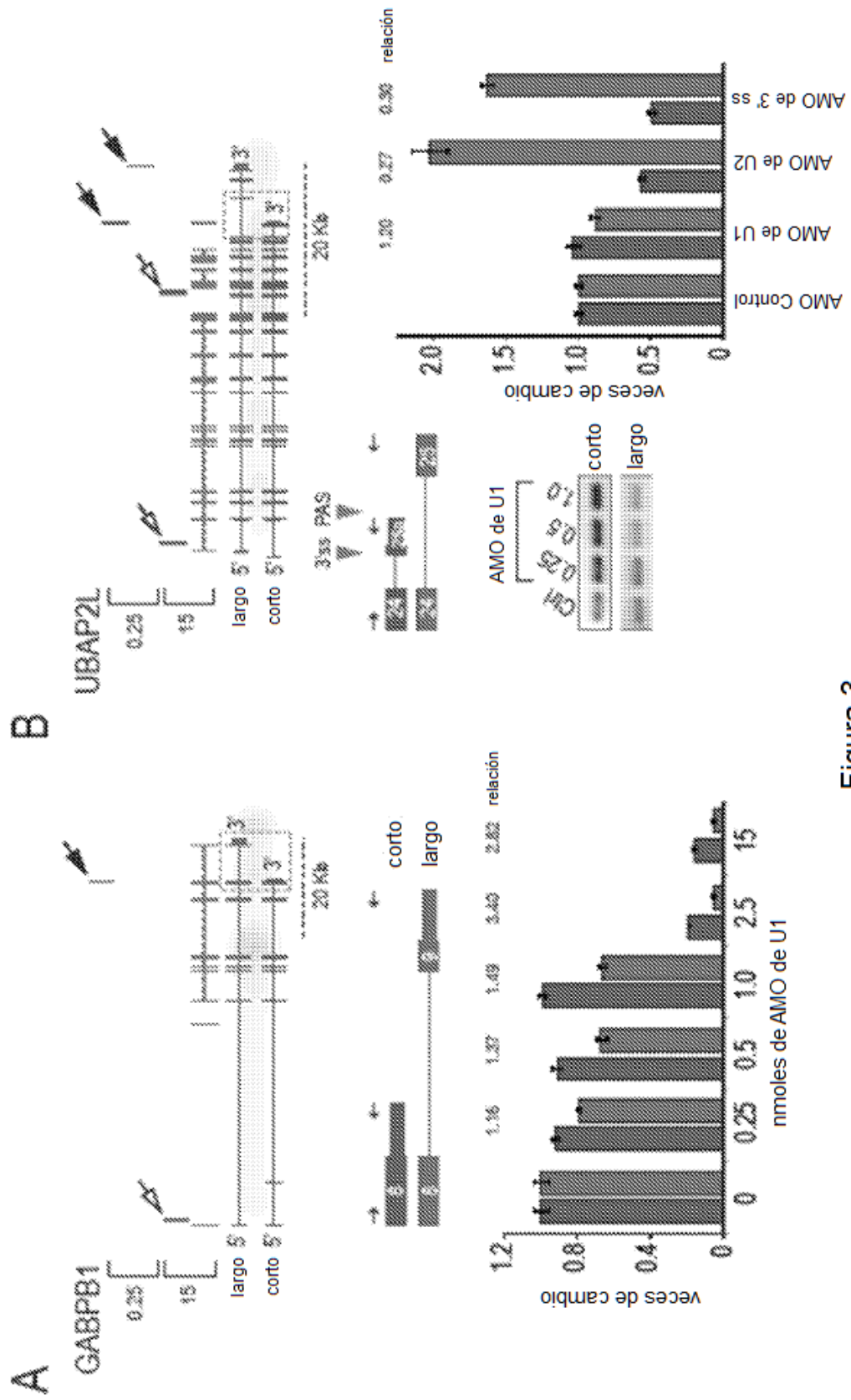


Figura 3

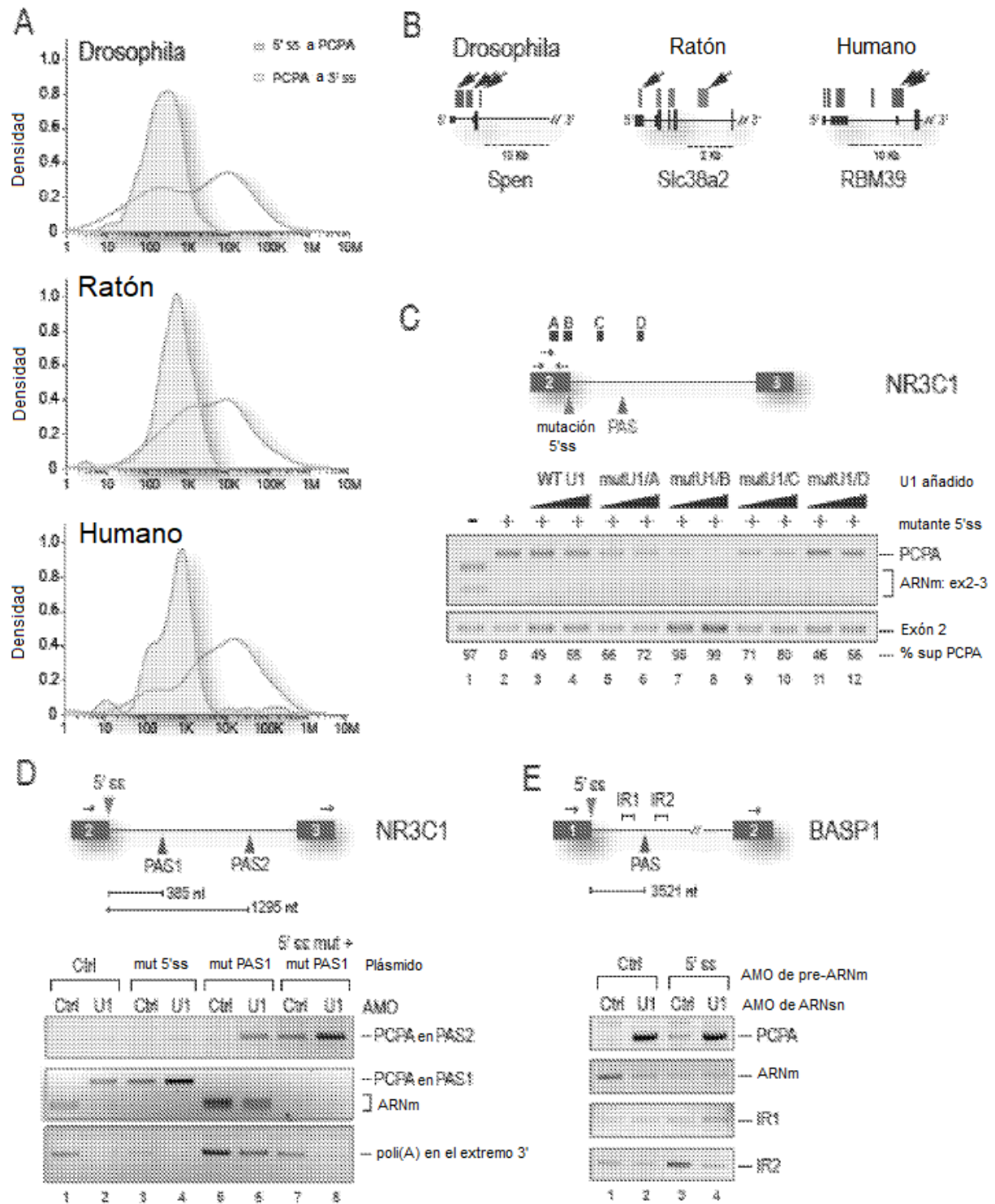


Figura 4

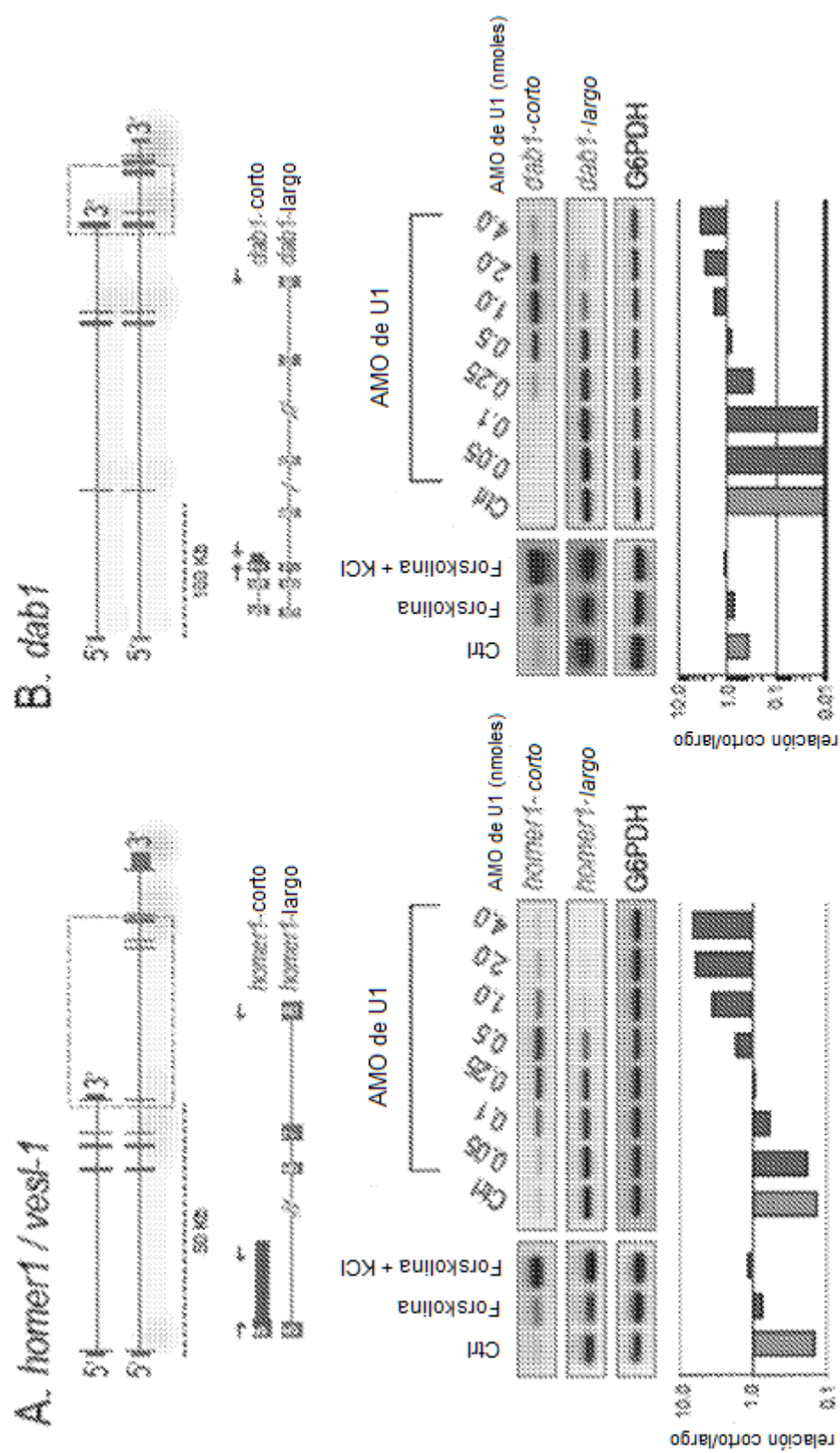


Figura 5

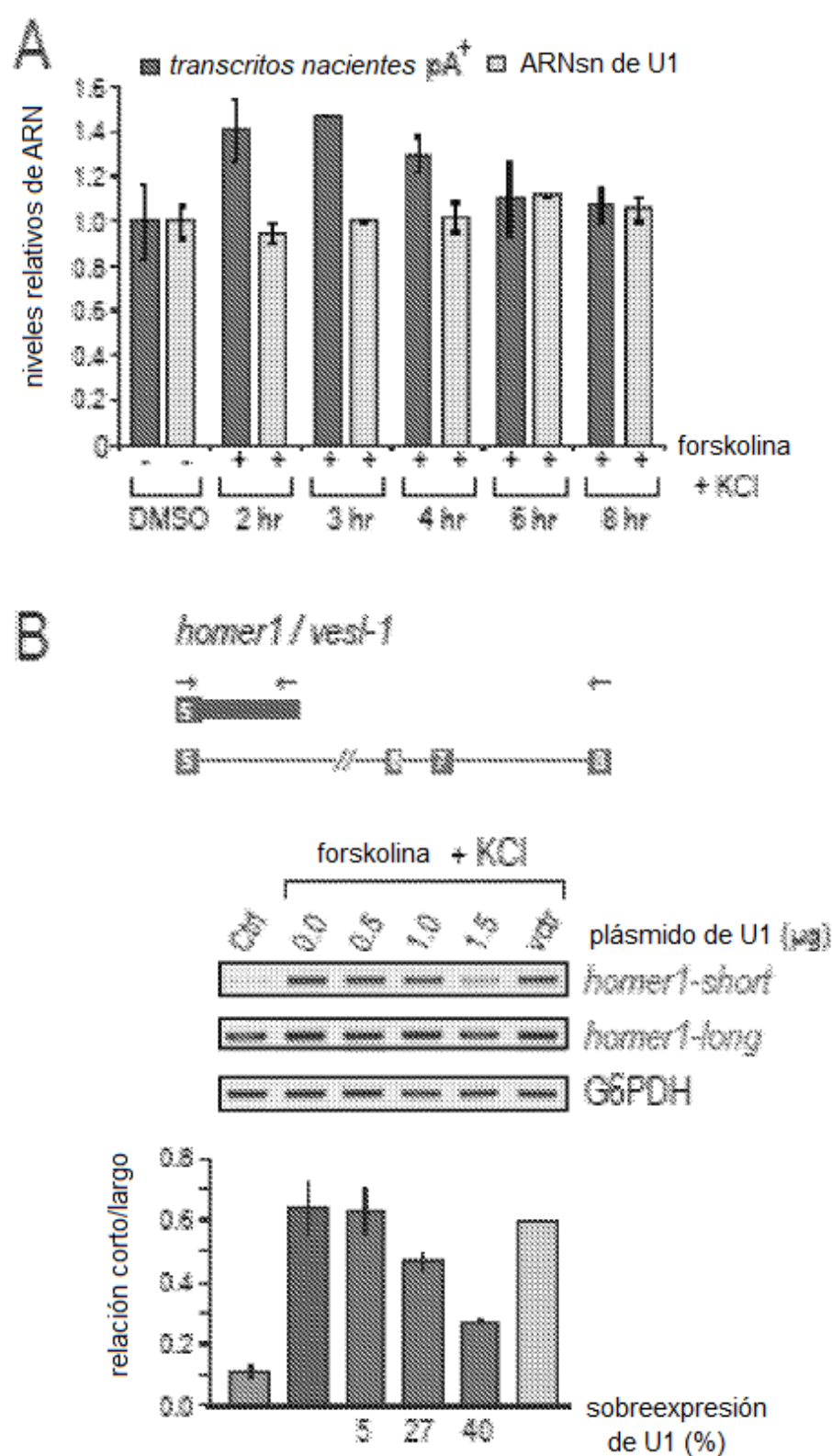


Figura 6

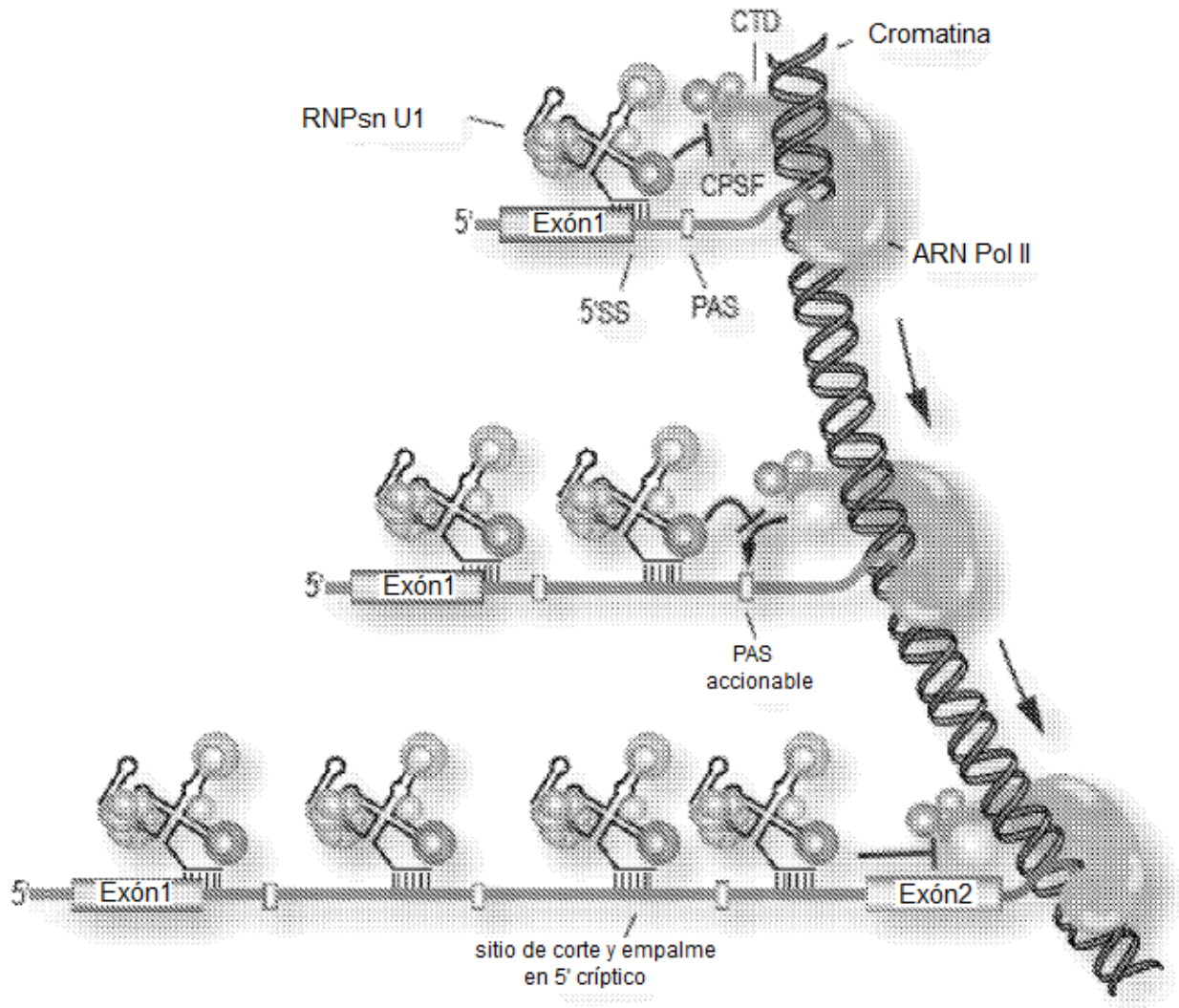
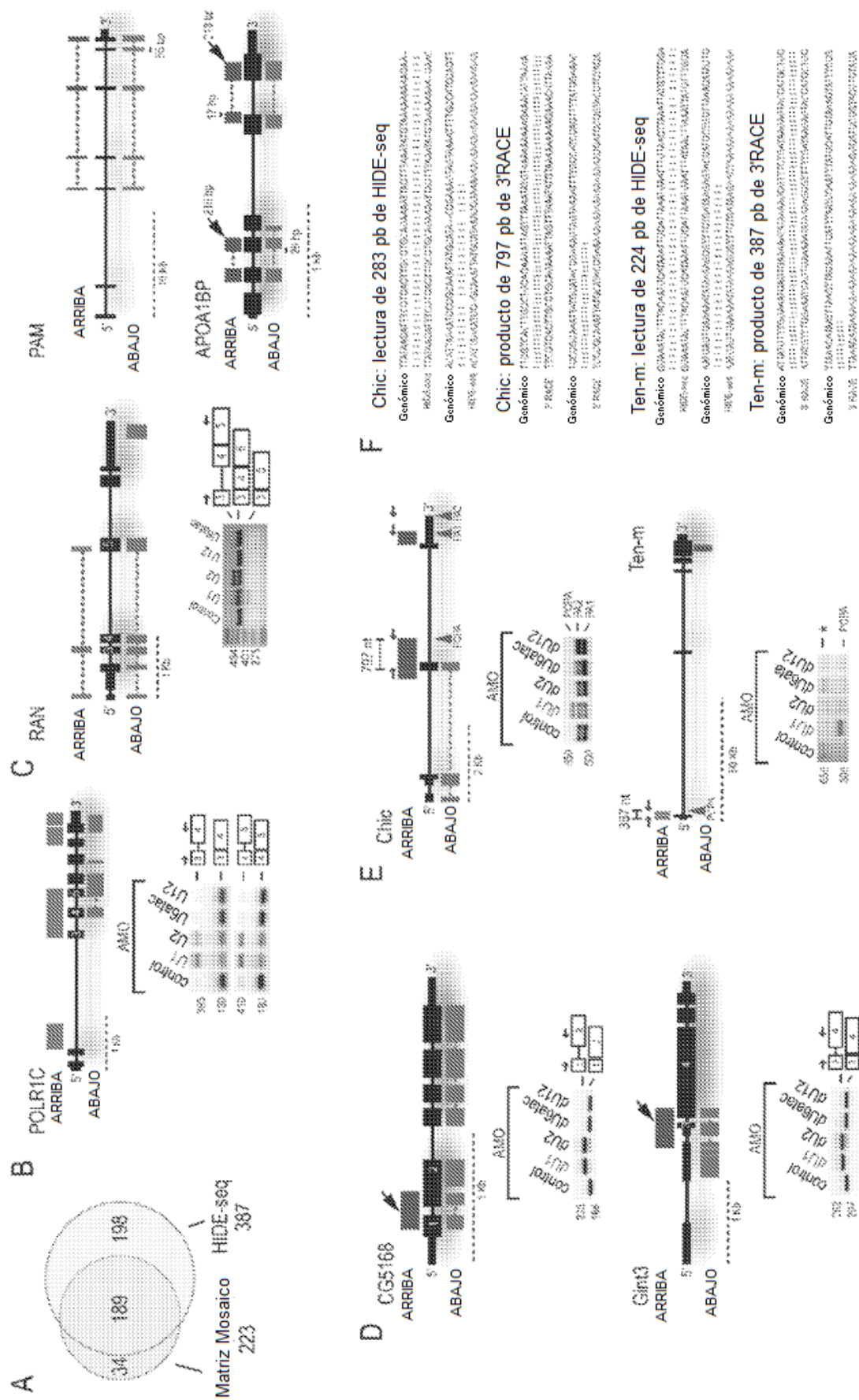
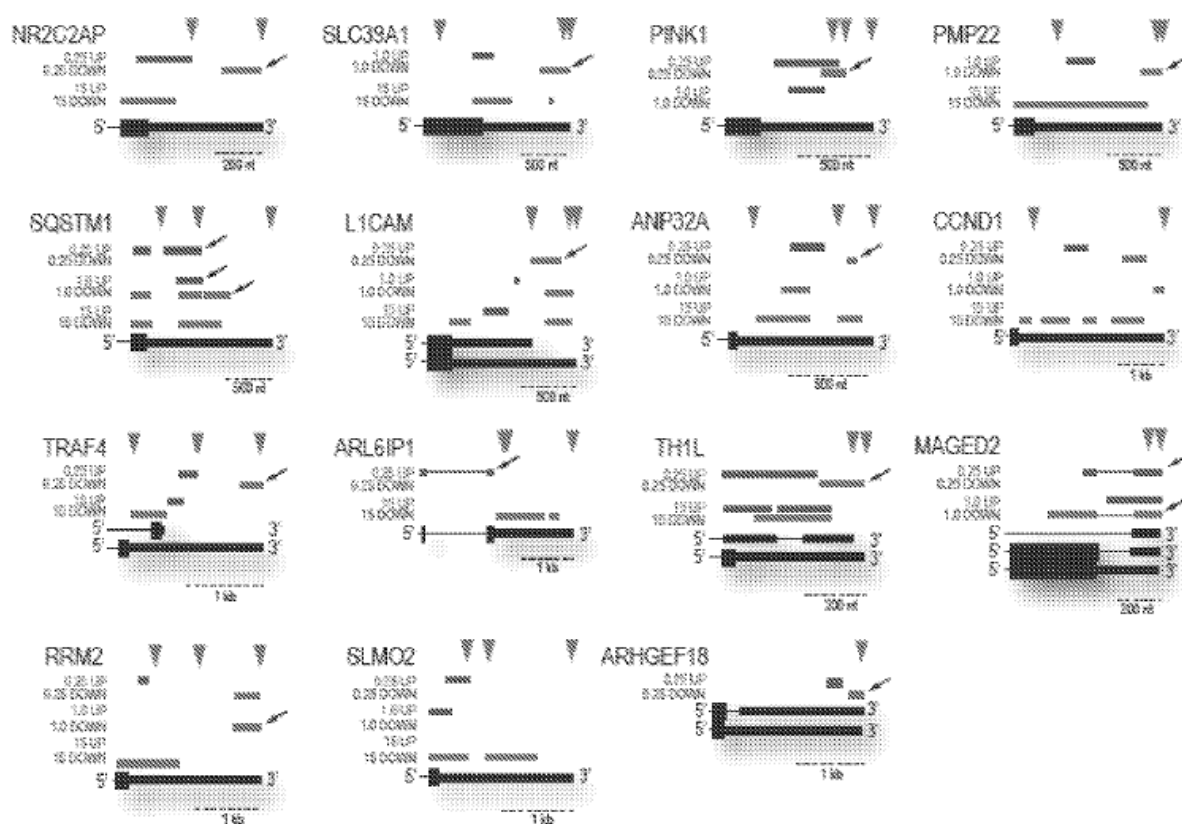


Figura 7



A



B

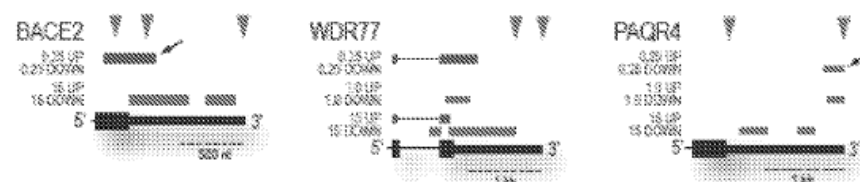


Figura 13

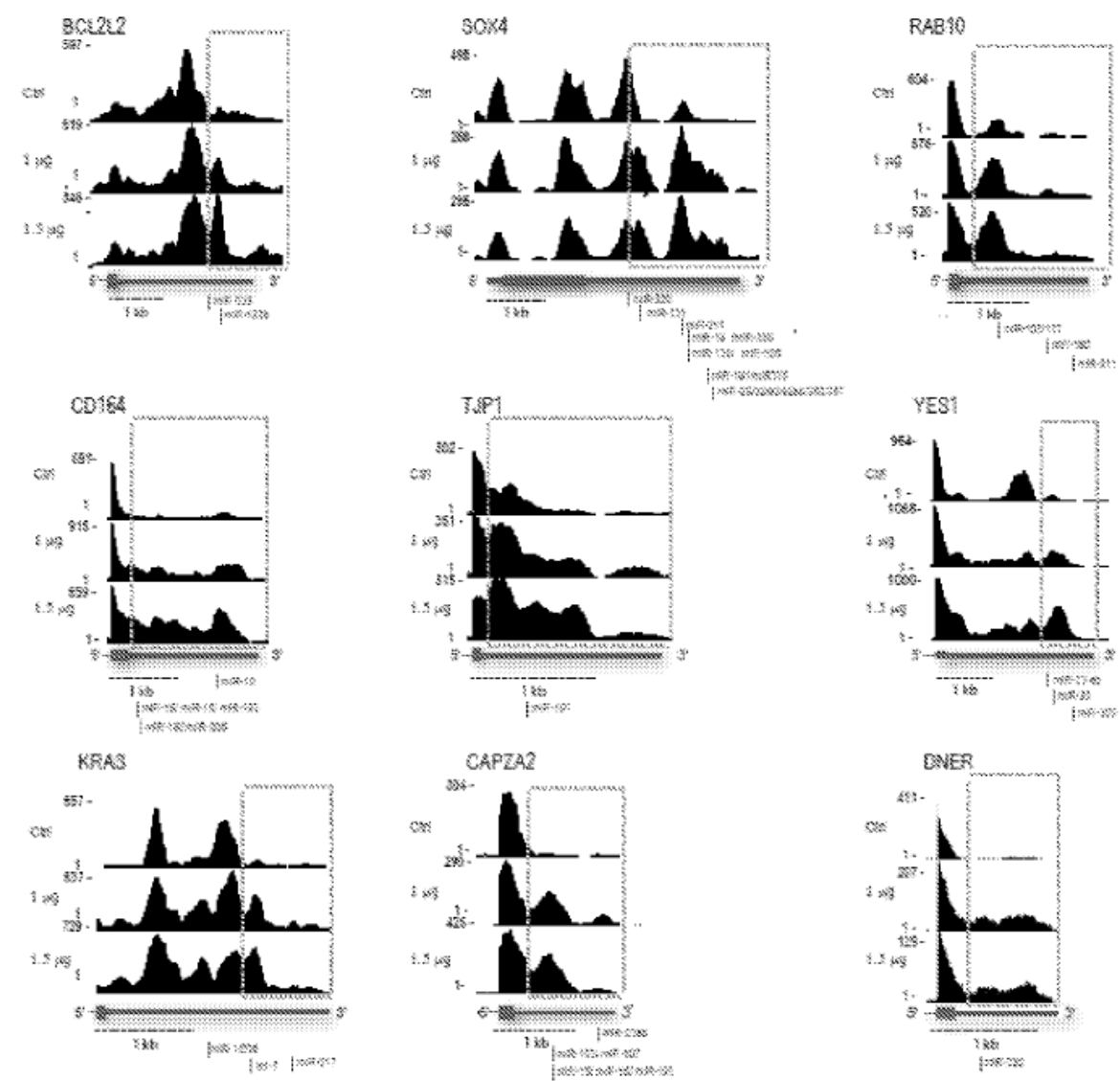


Figura 14

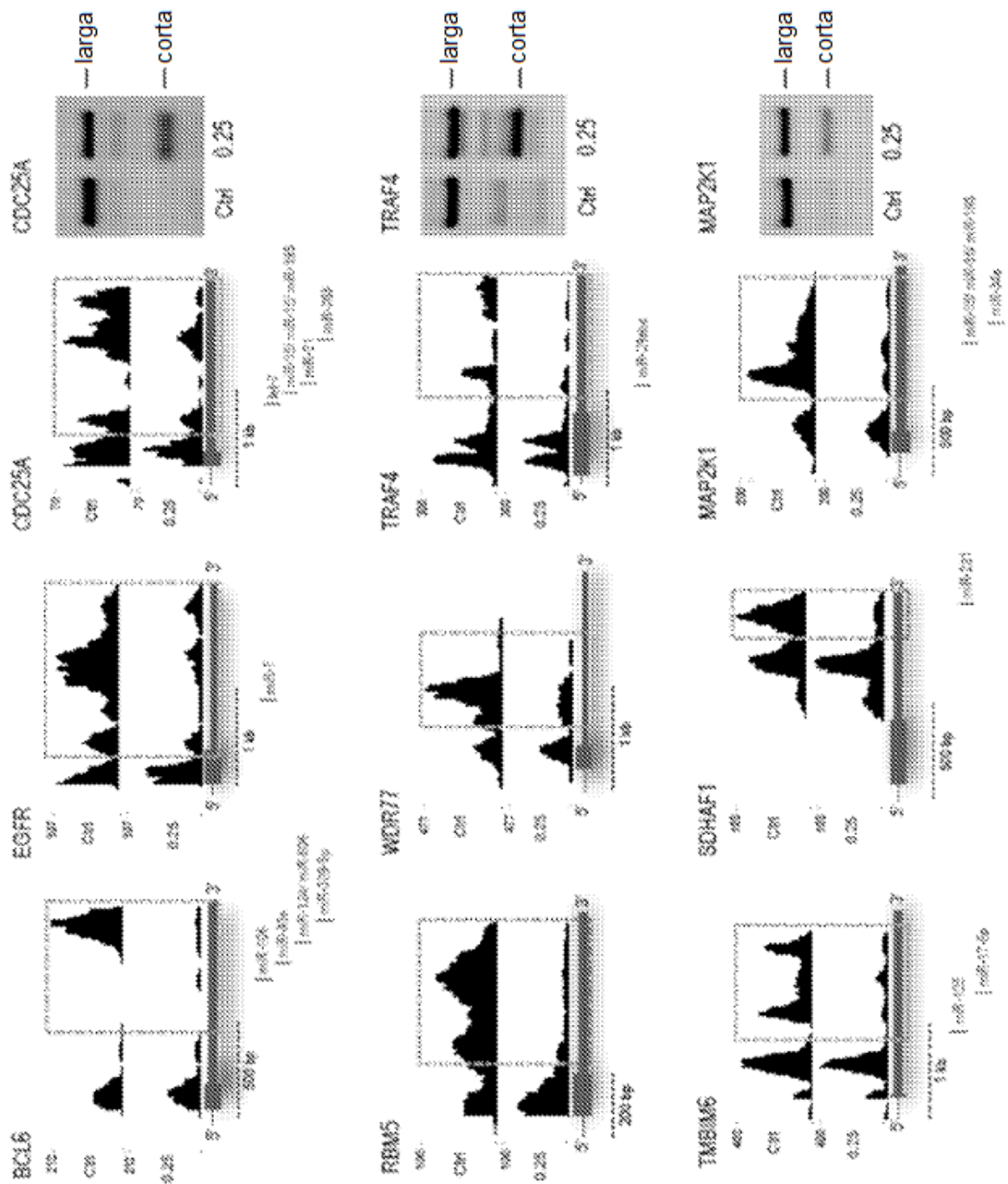


Figure 15

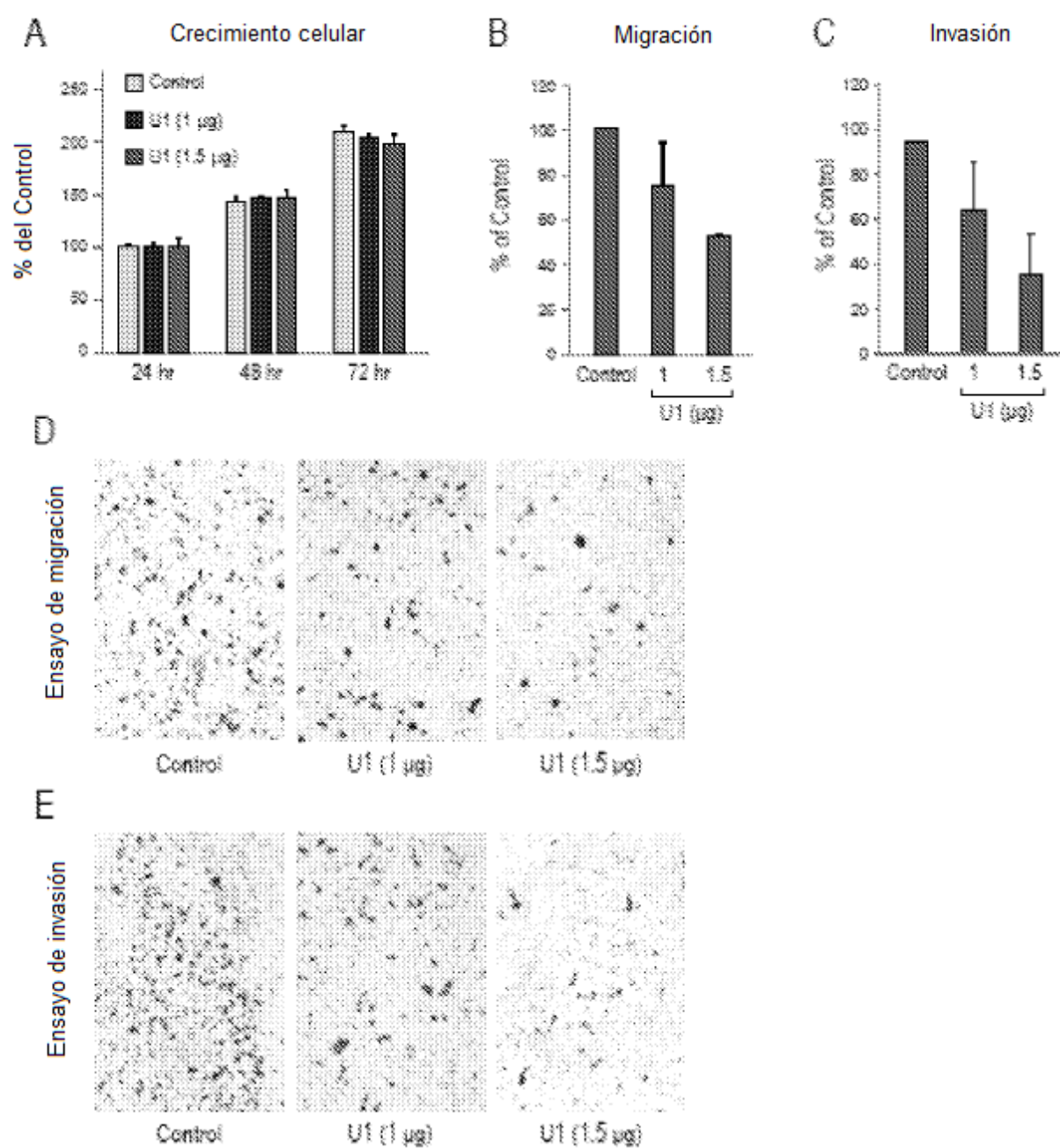


Figura 16

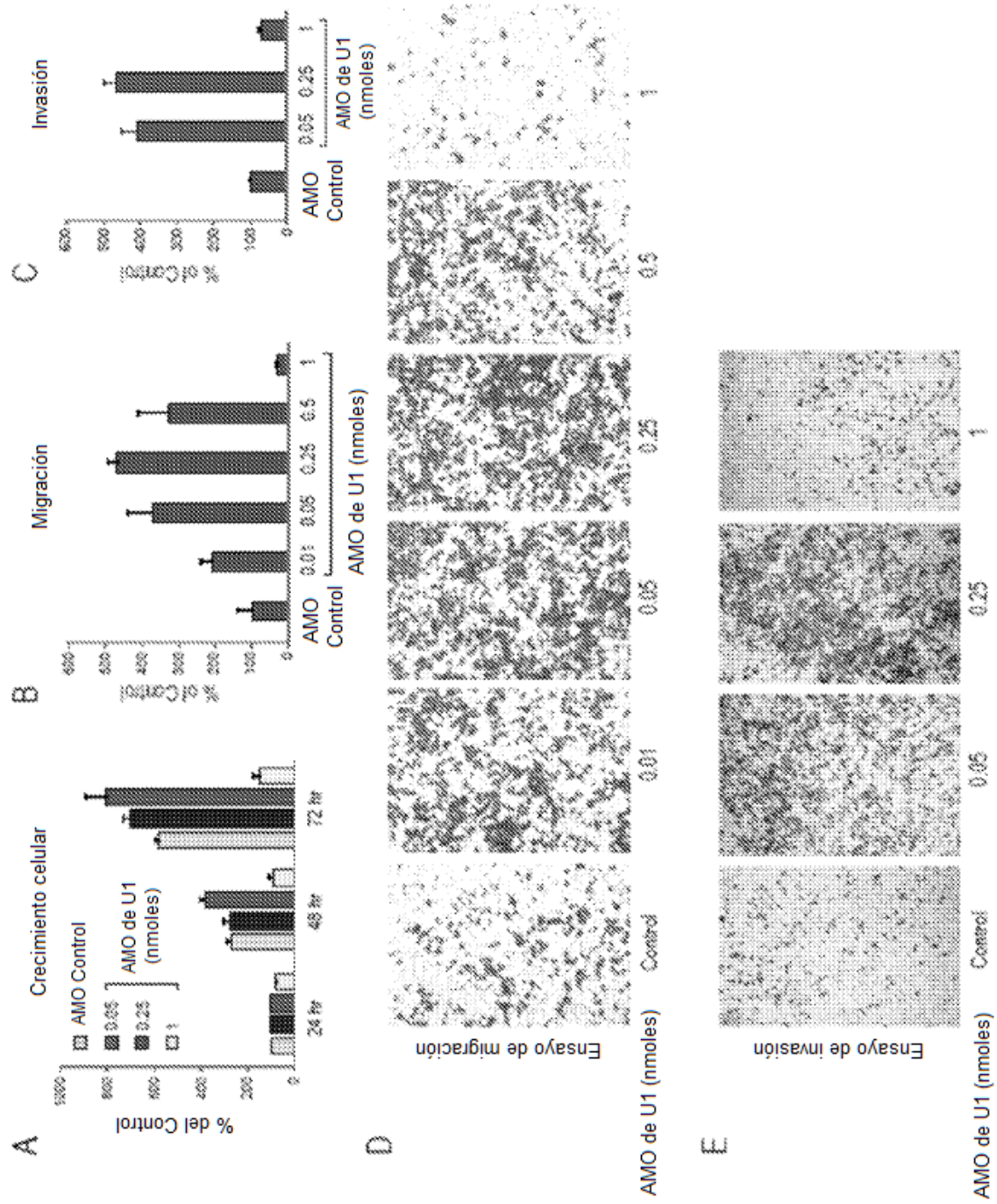


Figura 17

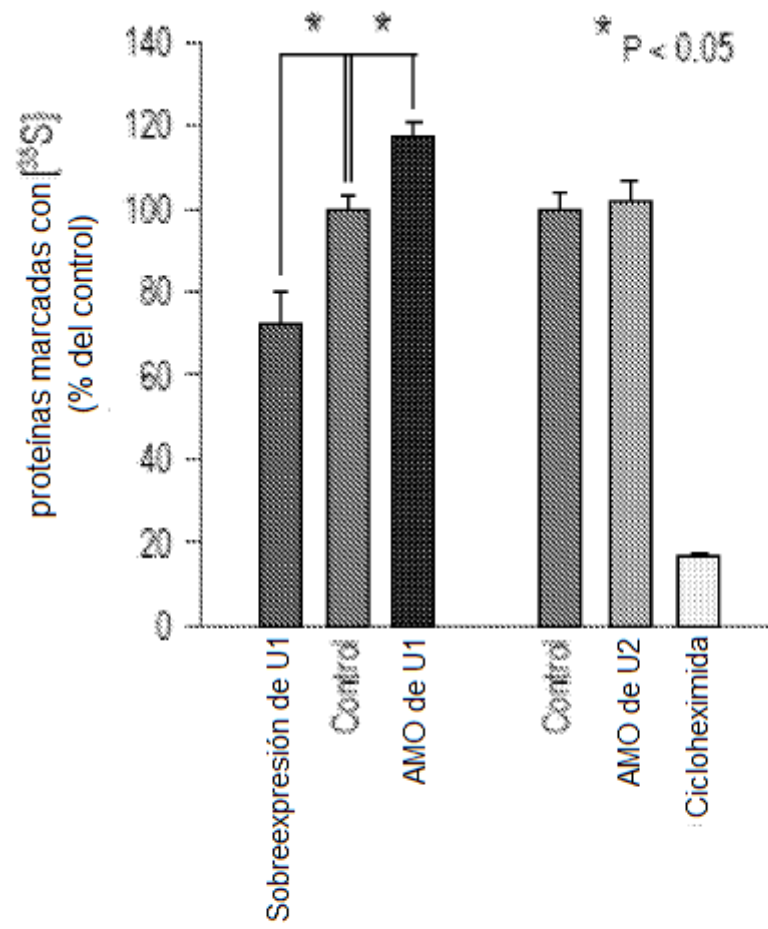


Figura 18

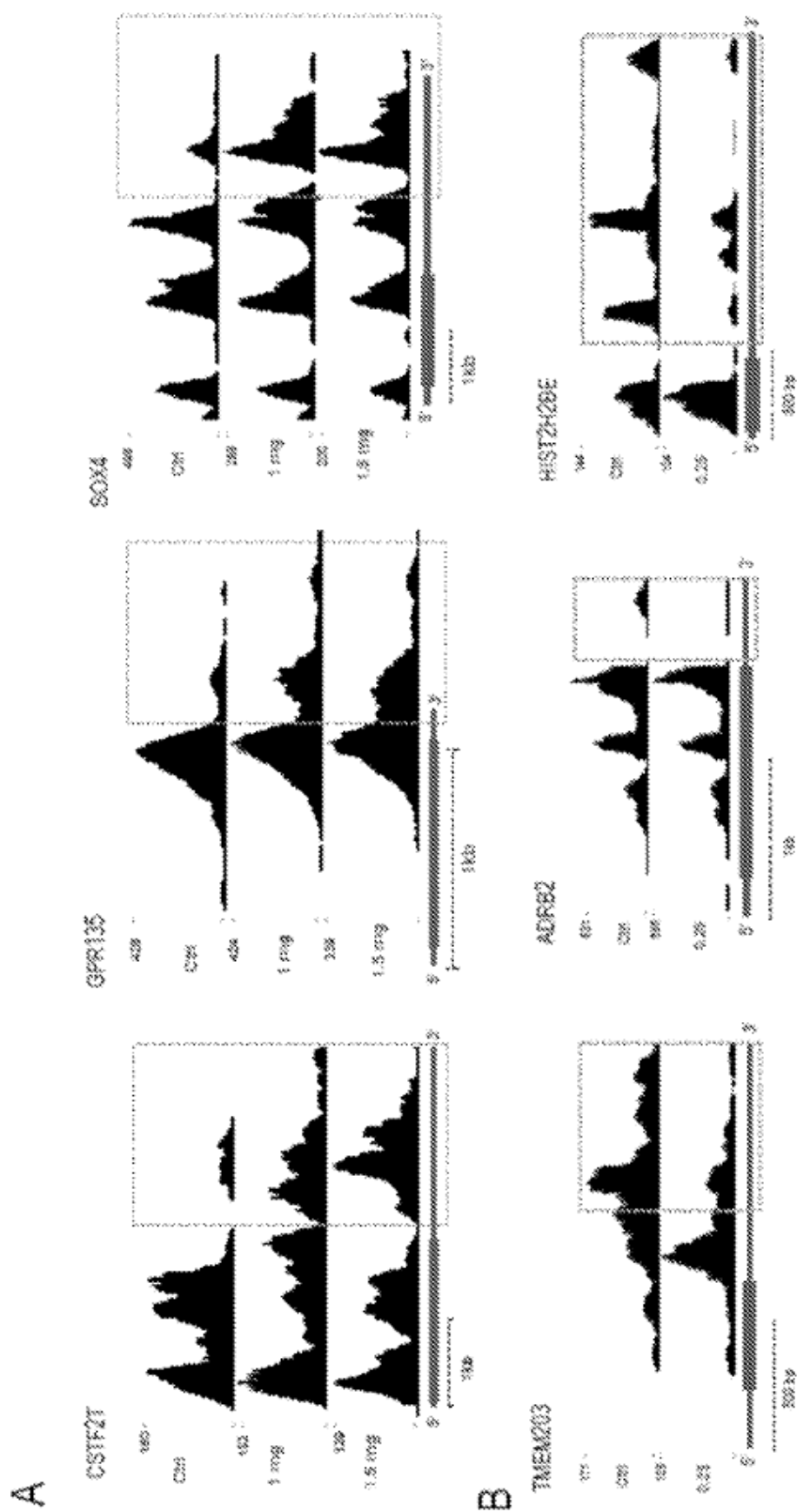


Figure 19

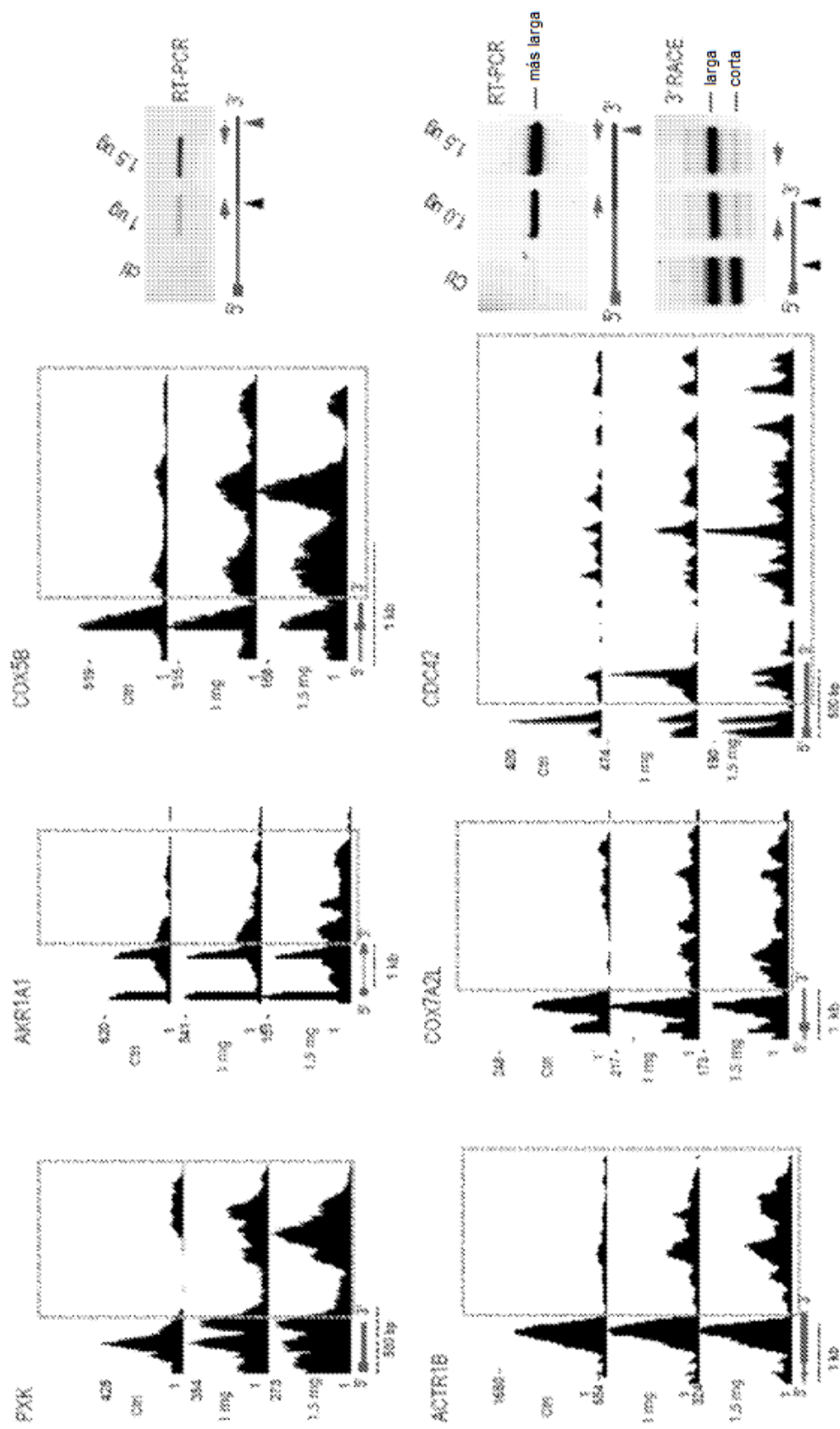


Figura 20



Menos agresivo HeLa Más agresivo

Oncogenes alargados por sobreexpresión de U1					Oncogenes acortados por disminución de U1	
AKAP9	ACSL3	ELK3	MOC	RPL22	ABL2	KIP1
ATRX	APC ^Δ	ERBB3 ^Δ	MITF ^Δ	SEDS	ARID1A	H0XC10 ^Δ
BCL10	ARHGEF12	EWSR1	MLL3	SDHD ^Δ	ASXL1	LYCAM
BCL2L2 ^Δ	ARID1A	EXT2	MN1 ^Δ	SEPT6	SAG1	LTBR
BRD3	ARID2	PBXW7	MNAT1	SETD2	BCL2 ^Δ	MDM2 ^Δ
CBFB	ATM	FOX ^Δ	MXI	SH3BP2	BCL6 ^Δ	MLL1
ETS1 ^Δ	BCL2 ^Δ	FOXP1	MYBL1	SLC45A3	CCDC5	MLLT6
FGFR1 ^Δ	BCOR	FUBP1	MYCL1 ^Δ	SRI	CCND1	MYH9
FIP1L1	BMPT1A	GAS1	NCOA1 ^Δ	TOF7L2 ^Δ	CCDC25A ^Δ	NUP98
HFE	BRAF	GNAQ	NCOA2	TFG	CDK12	RALB
KDM6A	BRCA2	GOLGA5	NN	TGFBR3	CHCHD7	RARA ^Δ
LRIG3	BTG1	GOPC	NONO	TNFAIP3 ^Δ	CRTC3	RPN1
MAML2	CBLB	GTF2H1	PAPAH1B2	TOP1	EGFR ^Δ	SDC4
MDM2 ^Δ	CCNC	H3F3A	PALB2	TOP2B	ERCC3	SET ^Δ
PRDM1	CD44	HERPUD1	PERM1	TRIM33	ETS2 ^Δ	SKP2 ^Δ
RAP2A	CD74	HEY1	PONA ^Δ	TRIP11	ETV4	TGFBR2 ^Δ
RASA1 ^Δ	CDK2	IGF1R ^Δ	PHF6	WH5C1	ETV6 ^Δ	TGFBR3
RBBP6	CDK6	LIFR	PICALM	WNT5A ^Δ	EWSR1	TIMP2
ZNF331	CDK7	LPP	PRKAR1A	WRN	FAM46C	TIMP3 ^Δ
ZNF384	CREB3L2	LYN ^Δ	RANBP17	WWTR1	FANCG	TRAF4
	DICER1 ^Δ	MAP2K4	RB1 ^Δ	YES1 ^Δ	FOSS	TSC1 ^Δ
	EGFR ^Δ				FYN ^Δ	WNT5A ^Δ
					HFE	

Figura 21

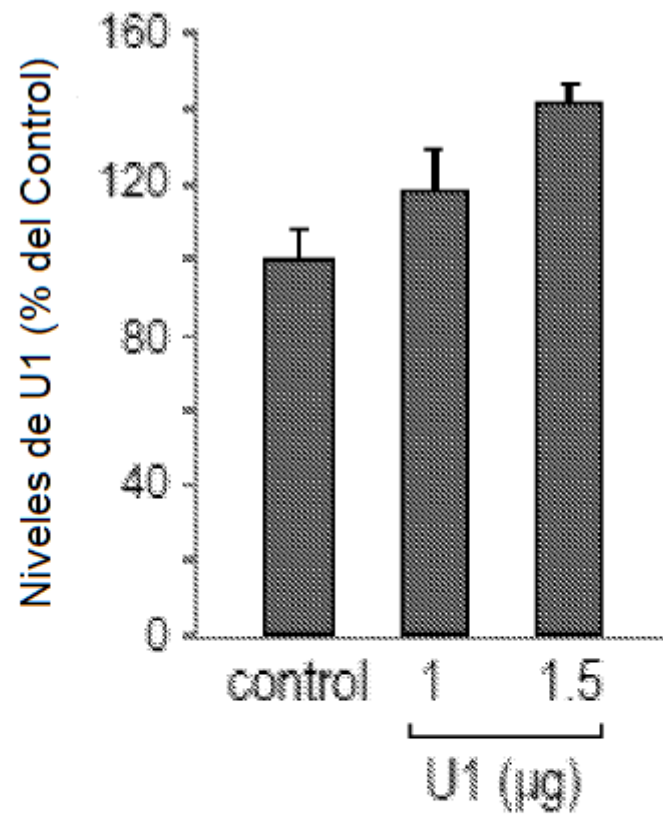


Figura 22

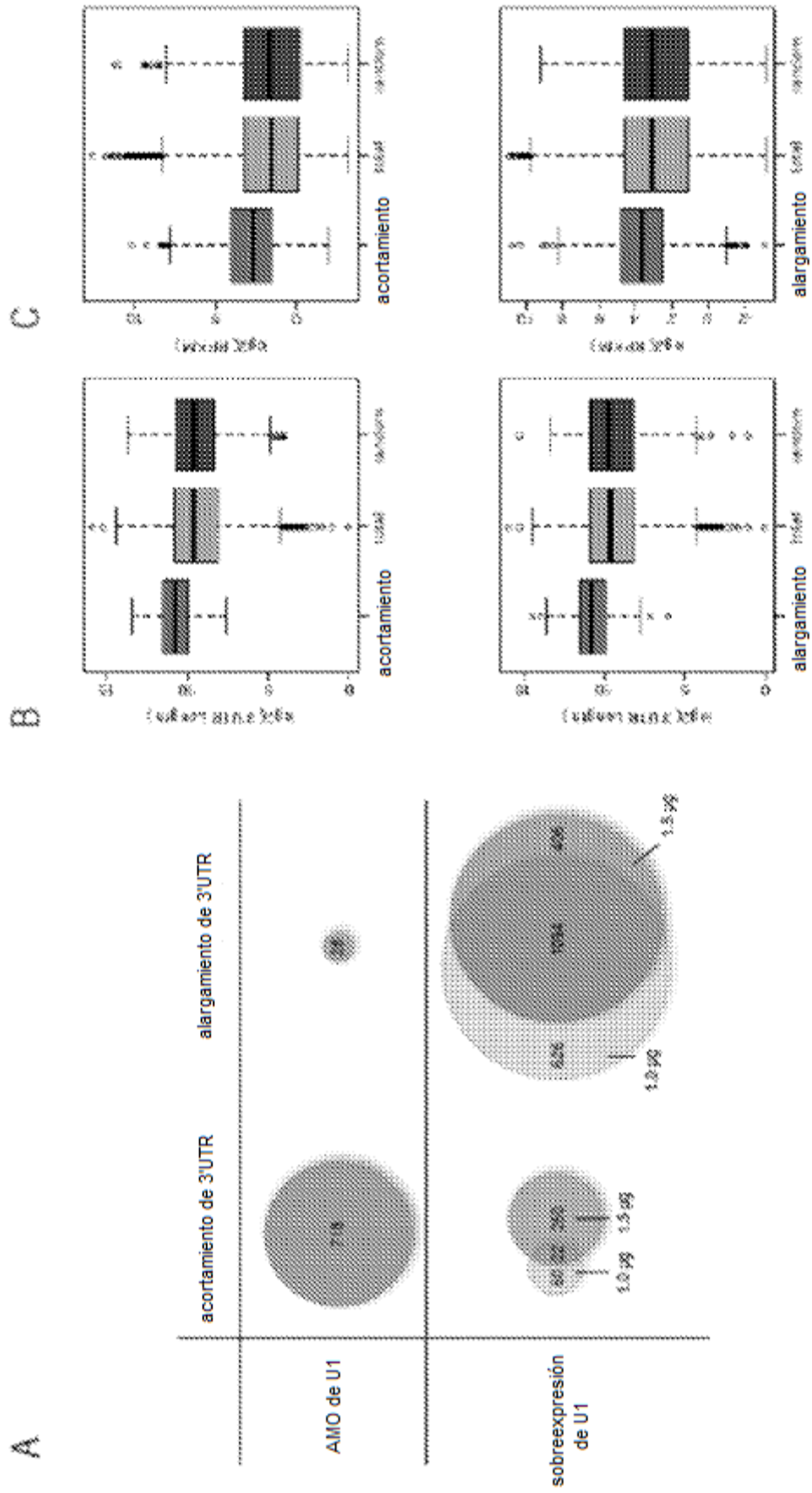


Figura 23

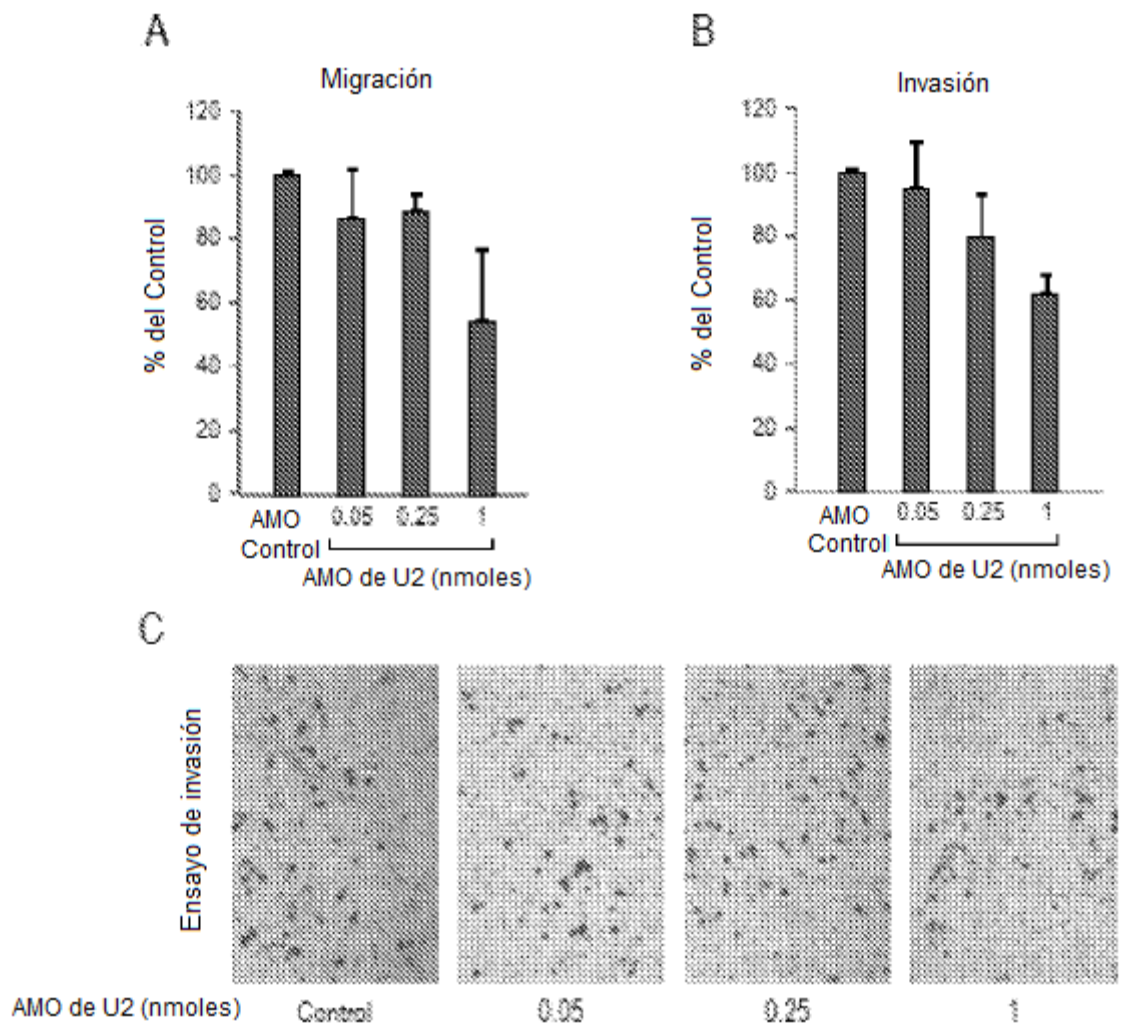


Figura 24