

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 812 242**

51 Int. Cl.:

C12N 15/00 (2006.01)

C12N 15/87 (2006.01)

A61K 48/00 (2006.01)

A01K 67/027 (2006.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.08.2014 E 17152034 (9)**

97 Fecha y número de publicación de la concesión europea: **10.06.2020 EP 3178928**

54 Título: **Ratón nuligénico para Pint que muestra un fenotipo asociado a envejecimiento prematuro**

30 Prioridad:

07.08.2013 US 201361863147 P

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

16.03.2021

73 Titular/es:

**REGENERON PHARMACEUTICALS, INC. (50.0%)
777 Old Saw Mill River Road
Tarrytown, NY 10591-6707, US y
PRESIDENT AND FELLOWS OF HARVARD
COLLEGE (50.0%)**

72 Inventor/es:

**LAI, KA-MAN VENUS;
GONG, GUOCHUN;
RINN, JOHN;
FRENDEWEY, DAVID y
VALENZUELA, DAVID M**

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 812 242 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

Ratón nuligénico para *Pint* que muestra un fenotipo asociado a envejecimiento prematuro

5 **Campo de la invención**

Los animales no humanos, células y tejidos y métodos para hacerlos, que comprenden una o más deficiencias en ARN no codificantes largos ("ARNlnc"). Los animales no humanos y métodos para hacerlos, que comprenden ARNlnc no funcionales o una inactivación (en inglés, knockout) de uno o más ARNlnc. Los animales no humanos modificados genéticamente que muestran fenotipos coherentes con un envejecimiento prematuro.

Antecedentes de la invención

Los ARN no codificantes largos (ARNlnc) y las subclases conocidas como ARN no codificantes intergénicos grandes (ARNlinc) comprenden aproximadamente 15.000 transcritos diversos en mamíferos que se asemejan a ARNm en su estructura, síntesis y el carácter cromatínico de sus genes. Las funciones o los fenotipos asociados con ARNlnc particulares no se conocen para la inmensa mayoría de ARNlnc identificados. Se piensa que algunos ARNlnc participan en el control transcripcional como activadores o represores transcripcionales en animales; otros pueden funcionar postraduccionalmente o mediante algún otro mecanismo. Por tanto, la capacidad para manipular ARNlnc puede proporcionar una herramienta para desarrollar fenotipos de interés que dependan de la identidad y función del ARNlnc. Existe una necesidad en la técnica de métodos y composiciones para manipular ARNlnc y una necesidad de generar fenotipos de animales no humanos a través de la manipulación de ARNlnc.

Sumario de la invención

La invención proporciona un ratón cuyo genoma comprende una inactivación de un locus del ARN no codificante largo (ARNlnc) *Pint* endógeno, en donde la inactivación (i) da como resultado que el ratón muestra un fenotipo asociado a envejecimiento prematuro que comprende (a) una tasa de crecimiento más lenta que la de un control de tipo silvestre; (b) un declive en la fuerza muscular; (c) fibrosis; (d) un contenido de grasa corporal más bajo que el del control de tipo silvestre; (e) una densidad mineral ósea del fémur y una masa ósea más baja que la del control de tipo silvestre; (f) una masa muscular disminuida en comparación con la del control de tipo silvestre; (g) una disminución en la longevidad media; (h) cifolordosis; o (i) una combinación de cualquiera de (a)-(h); y (ii) comprende una delección de la mayor parte de la secuencia que codifica el ARNlnc *Pint* que comienza en un segundo exón del locus del ARNlnc *Pint* endógeno y que da como resultado una mutación de pérdida de función de *Pint*.

La invención también proporciona:

- una célula de ratón, tejido o embrión cuyo genoma comprende una inactivación de un locus del ARN no codificante largo (ARNlnc) *Pint* endógeno, en donde la inactivación comprende una delección de la mayor parte de la secuencia codificante del ARNlnc *Pint* que comienza en un segundo exón del locus *Pint* endógeno y da como resultado una mutación de pérdida de función de *Pint*; y
- un método para hacer un ratón de la invención, que comprende: (1)

poner en contacto una célula madre embrionaria (ES; del inglés, embryonic stem) de ratón con un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3'; en donde el constructo de direccionamiento sufre recombinación homóloga con el locus del ARNlnc *Pint* endógeno de ratón en un genoma de la célula ES de ratón para formar una célula ES de ratón modificada; (2) introducir la célula ES de ratón modificada en un embrión de ratón hospedador; y (3) gestar el embrión de ratón hospedador en una madre ratón subrogada, en donde la madre ratón subrogada produce una progenie que comprende la inactivación del locus del ARNlnc *Pint* endógeno.

En el presente documento se describen animales no humanos, células, tejidos y embriones que comprenden ARN no codificantes largos no funcionales (ARNlnc), que incluyen, pero no se limitan a las inactivaciones de uno o más ARNlnc. Se describen en el presente documento métodos y composiciones para manipular la expresión de ARNlnc. Las composiciones de direccionamiento dirigidas a modificar o inactivar ARNlnc también se describen en el presente documento. En el presente documento también se describen animales no humanos, células y tejidos que muestran un fenotipo asociado con la no función de uno o más ARNlnc.

En un aspecto descrito en el presente documento, se proporciona un animal no humano que comprende en su genoma al menos un locus de ARNlnc modificado, en donde el locus de ARNlnc modificado comprende una mutación de pérdida de función en una secuencia de ácido nucleico que codifica un ARNlnc.

En un aspecto descrito en el presente documento, el ARNlnc es un ARN no codificante intergénico largo (ARNlinc).

En un aspecto descrito en el presente documento, la mutación de pérdida de función se caracteriza por una alteración

o una inactivación de al menos una función de ARNlnc.

5 En un aspecto descrito en el presente documento, el locus de ARNlnc modificado comprende una delección de uno o más exones que codifican el ARNlnc o una porción del mismo. En un aspecto descrito en el presente documento, la alteración o inactivación comprende una delección de uno o más exones dentro del locus de ARNlnc que comienza en un segundo exón de un locus de ARNlnc; una delección de uno o más exones dentro del locus de ARNlnc que comienza en un primer exón de un locus de ARNlnc; o delección de una región codificante de ARN completa de un locus de ARNlnc.

10 En un aspecto descrito en el presente documento, la alteración o inactivación comprende un reemplazo de un locus de ARNlnc o una porción del mismo con un inserto de ácido nucleico. En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende una primera secuencia de nucleótidos que codifica un indicador. En algunos de tales casos, la primera secuencia de nucleótidos está ligada operativamente a un promotor que impulsa la expresión del indicador. En un aspecto descrito en el presente documento, la primera secuencia de nucleótidos que codifica el indicador se posiciona en un locus de ARNlnc en unión operativa con un promotor de ARNlnc endógeno, en donde el promotor de ARNlnc endógeno impulsa la expresión de la secuencia de nucleótidos. En tales casos, la expresión de la secuencia de ácidos nucleicos sigue un patrón de expresión del ARNlnc. En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende una secuencia consenso Kozak. En un aspecto específico descrito en el presente documento, la primera secuencia de nucleótidos del inserto de ácido nucleico comprende una secuencia consenso Kozak.

25 En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende adicionalmente una segunda secuencia de nucleótidos que codifica un marcador seleccionable, en donde la segunda secuencia de nucleótidos está ligada operativamente a un promotor.

En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende sitios de recombinación específica de sitio que flanquean un segmento que codifica el indicador y/o un segmento que codifica un marcador seleccionable.

30 En diversos aspectos descritos en el presente documento, el reemplazo del locus de ARNlnc o porción del mismo comprende el reemplazo de uno o más exones dentro de un locus de ARNlnc que comienza en el segundo exón del locus de ARNlnc con el inserto de ácido nucleico; el reemplazo de uno o más exones dentro de un locus de ARNlnc que comienza en el primer exón del locus de ARNlnc con el inserto de ácido nucleico; o el reemplazo de la región codificante de ARN completa de un locus de ARNlnc con el inserto de ácido nucleico.

35 En un aspecto descrito en el presente documento, un animal no humano proporcionado en el presente documento se caracteriza por tener uno o más de los siguientes fenotipos: (a) un fenotipo asociado a envejecimiento prematuro; (b) letalidad perinatal; (c) un defecto en el desarrollo pulmonar; (d) una malformación morfológica en la cola y en las extremidades traseras; (e) una pérdida de masa muscular en uno o más tejidos; o (f) una combinación de los mismos de cualquiera de (a)-(e).

45 En un aspecto descrito en el presente documento, un animal no humano proporcionado en el presente documento comprende una alteración o inactivación del ARNlnc *Pint* y el animal no humano se caracteriza por un fenotipo asociado a envejecimiento prematuro que comprende: (a) una tasa de crecimiento más lenta que la de un control de tipo silvestre; (b) un declive en la fuerza muscular; (c) fibrosis; (d) un contenido de grasa corporal más bajo que el del control de tipo silvestre; (e) una densidad mineral ósea del fémur y una masa ósea más baja que la del control de tipo silvestre; (f) una masa muscular disminuida en comparación con la del control de tipo silvestre; (g) una disminución en la longevidad media; (h) cifolordosis; (i) atrofia orgánica; o (j) una combinación de los mismos de cualquiera de (a)-(i).

50 En un aspecto descrito en el presente documento, un animal no humano proporcionado en el presente documento muestra un defecto en el desarrollo cerebral. En algunos de tales casos, el ARNlnc es *Pantr2*, *Kantr*, *Peril*, *Celrr*, *Pantr1*, *Crnde*, *lincenc1*, *Pint*, *lincppara* o *Tug1*.

55 En diversos aspectos descritos en el presente documento, el animal no humano es un mamífero. En diversos aspectos descritos en el presente documento, el mamífero es un roedor, por ejemplo, un ratón, una rata o un hámster. En diversos aspectos descritos en el presente documento, el mamífero es una especie ovina, bovina o porcina.

60 En un aspecto descrito en el presente documento, se proporciona un animal no humano modificado genéticamente, en donde la modificación genética da como resultado una pérdida de función de un ARNlnc.

En un aspecto descrito en el presente documento, se proporciona un animal no humano modificado genéticamente, en donde la modificación genética comprende una alteración o una inactivación de uno o más ARNlnc.

65 En un aspecto descrito en el presente documento, la modificación genética comprende una alteración o inactivación de al menos dos ARNlnc. En un aspecto descrito en el presente documento, la modificación genética comprende una alteración o una inactivación de al menos tres, cuatro, cinco o seis ARNlnc.

- En un aspecto descrito en el presente documento, la modificación genética comprende una alteración o inactivación de uno o más ARNlnc y dentro del locus de ARNlnc un gen que codifica una fracción detectable (indicador) unido operativamente a un promotor de un ARNlnc alterado o inactivado. En un aspecto descrito en el presente documento, el gen que codifica la fracción detectable (indicador) se selecciona de lacZ (que codifica β -galactosidasa), GFP, eGFP, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrina, Venus, YPet, proteína fluorescente amarilla potenciada (EYFP; del inglés, enhanced yellow fluorescent protein), Esmeralda, CyPet, proteína fluorescente cian (CFP; del inglés, cyan fluorescent protein), Cerulean, T-Zafiro, luciferasa, fosfatasa alcalina y una combinación de los mismos.
- En un aspecto descrito en el presente documento, la modificación genética comprende una alteración o inactivación de uno o más ARNlnc y dentro del locus de ARNlnc un gen que codifica una fracción detectable (indicador) unido operativamente a un promotor que impulsa la expresión de la fracción detectable.
- En diversos aspectos descritos en el presente documento, la fracción detectable incluye cualquier gen indicador conocido en la técnica.
- En un aspecto descrito en el presente documento, se proporciona un animal no humano modificado genéticamente, en donde la modificación genética comprende una alteración de una secuencia de ácido nucleico de ARNlnc que codifica un ARNlnc seleccionado del grupo que consiste en *HOTAIR*, *HOTTIP*, *Hoxallos* (denominado anteriormente *HoxA11as*), *Pantr1* (denominado anteriormente *lincRNA-Brn1-a*), *Pantr2* (denominado anteriormente *lincRNA-Brn1-b*), *Ptgs2os2* (denominado anteriormente *lincRNA-Cox2*), *Eldr* (denominado anteriormente *Fabl* y *lincRNA-Egfr*), *Lincenc1* (denominado anteriormente *lincRNA-Enc1*), *Mannr* (denominado anteriormente *lincRNA-Evil*), *Fendrr* (denominado anteriormente *lincRNA Foxf1*), *Halr1* (denominado anteriormente *Haunt* y *lincRNA-HoxA1*), *Haglrl* (denominado anteriormente *lincRNA-HoxD3*), *Celrr* (denominado anteriormente *Celr* y *lincRNA-Insig2*), *Crnde* (denominado anteriormente *lincRNA-Irx5*), *Kantr* (denominado anteriormente *Spasm* y *lincRNA-Jaridlc*), *Pint* (denominado anteriormente *line-Pint* y *lincRNA-Mkln1*), *Trp53cor1* (denominado anteriormente *lincRNA-p21*), *lincppara* (denominado anteriormente *lincRNA-Ppara*), *Peril* (denominado anteriormente *lincRNA-Sox2*), *Tug1* (denominado anteriormente *lincRNA-Tug1*) y una combinación de los mismos.
- En un aspecto descrito en el presente documento, se proporciona un animal no humano modificado genéticamente, en donde la modificación genética comprende una alteración o inactivación de *Pint* (denominado anteriormente *line-Pint* y *lincRNA-Mkln1*).
- En un aspecto descrito en el presente documento, se proporciona un locus de ácido nucleico, que comprende una alteración de un ARNlnc. En un aspecto descrito en el presente documento, la alteración comprende una inactivación del ARNlnc. En un aspecto descrito en el presente documento, la alteración comprende una colocación de un gen que codifica una fracción detectable ligada operativamente a un promotor del ARNlnc. En un aspecto descrito en el presente documento, la alteración comprende una inactivación del ARNlnc y una colocación de un gen que codifica una fracción detectable en unión operativa con el promotor del ARNlnc.
- En un aspecto descrito en el presente documento, se proporciona un constructo de ácido nucleico, que comprende al menos una secuencia de direccionamiento que dirige el constructo hacia un locus que comprende un ARNlnc, en donde el constructo es capaz de alterar la transcripción del ARNlnc, inactivando el ARNlnc o reemplazando el ARNlnc.
- En un aspecto descrito en el presente documento, el constructo de ácido nucleico comprende adicionalmente una fracción detectable (con o sin promotor añadido que impulsa la expresión de la fracción detectable). En un aspecto descrito en el presente documento, el constructo de ácido nucleico comprende adicionalmente un gen marcador seleccionable impulsado por un promotor. En un aspecto descrito en el presente documento, el constructo de ácido nucleico comprende tanto una fracción detectable (con o sin su propio promotor) y un gen marcador seleccionable impulsado por un promotor. En un aspecto descrito en el presente documento, el marcador seleccionable y/o la fracción detectable está(n) flanqueado(s) en dirección 5' y en dirección 3' con sitios de recombinación específica de sitio que dirigen una escisión de la fracción detectable y/o el marcador seleccionable.
- En un aspecto descrito en el presente documento, se proporciona un vector de direccionamiento. En un aspecto, el vector de direccionamiento comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3' que pueden someterse a recombinación homóloga con un locus de ARNlnc de interés. En un aspecto descrito en el presente documento, el inserto de ácido nucleico del vector de direccionamiento comprende una primera secuencia de nucleótidos que codifica un indicador. En un aspecto descrito en el presente documento, tras la recombinación homóloga con el locus de ARNlnc de interés, la primera secuencia de ácido nucleico que codifica el indicador está ligada operativamente a un promotor endógeno que impulsa la expresión de un ARNlnc en el locus de ARNlnc. En un aspecto descrito en el presente documento, la primera y/o segunda secuencia de ácido nucleico del inserto de ácido nucleico del vector de direccionamiento comprende(n) adicionalmente una secuencia consenso Kozak. En un aspecto descrito en el presente documento, el vector de direccionamiento comprende adicionalmente un promotor que impulsa la expresión del promotor.

- En un aspecto descrito en el presente documento, el inserto de ácido nucleico del vector de direccionamiento comprende adicionalmente una segunda secuencia de nucleótidos que codifica un marcador seleccionable, en donde el segundo ácido nucleico está unido operativamente a un promotor. En un aspecto descrito en el presente documento, el vector de direccionamiento comprende adicionalmente sitios de recombinación específica de sitio que flanquean un segmento que codifica el indicador y/o un segmento que codifica el ácido nucleico del marcador seleccionable.
- En un aspecto descrito en el presente documento, se proporciona un fenotipo que muestra un fenotipo asociado a envejecimiento prematuro, en donde el animal no humano comprende una modificación que hace que *Pint* sea no funcional. En un aspecto descrito en el presente documento, la modificación es una alteración de una secuencia codificante de ARN de un locus *Pint*. En un aspecto descrito en el presente documento, la modificación es una delección de una secuencia codificante de ARN completa de un locus *Pint*. En un aspecto descrito en el presente documento, la modificación comprende una inserción de un vector de direccionamiento en un locus *Pint*, de manera que el animal ya no hace un *Pint* funcional.
- En un aspecto descrito en el presente documento, la modificación comprende adicionalmente en el locus *Pint* una fracción detectable (por ejemplo, un gen indicador con o sin su propio indicador) y opcionalmente un marcador seleccionable. En un aspecto descrito en el presente documento, la fracción seleccionable y/o el marcador seleccionable está(n) flanqueado(s) en dirección 5' y en dirección 3' con sitios de recombinación específica de sitio que dirigen una escisión de la fracción detectable y/o el marcador seleccionable. En un aspecto descrito en el presente documento, el animal no humano comprende adicionalmente una recombinasa específica de sitio inducible que es compatible con los sitios de recombinasa específica de sitio.
- En un aspecto descrito en el presente documento, se proporciona una célula, tejido o embrión de un animal no humano, en donde la célula o tejido carece de una secuencia de ácido nucleico que codifica un ARNlinc funcional seleccionado del grupo que consiste en *HOTAIR*, *HOTTIP*, *Hoxallos* (denominado anteriormente *HoxA11as*), *Pantr1* (denominado anteriormente *lincRNA-Brn1-a*), *Pantr2* (denominado anteriormente *lin-cRNA-Brn1-b*), *Ptgs2os2* (denominado anteriormente *lincRNA-Cox2*), *Eldr* (denominado anteriormente *Fabl* y *lincRNA-Egfr*), *Lincenc1* (denominado anteriormente *lincRNA-Enc1*), *Mannr* (denominado anteriormente *lincRNA-Evil*), *Fendrr* (denominado anteriormente *lincRNA-Foxf1*), *Halr1* (denominado anteriormente *Haunt* y *lincRNA-HoxA1*), *Haglr* (denominado anteriormente *Mdgt* y *lincRNA-HoxD3*), *Celrr* (denominado anteriormente *Celr* y *lincRNA-Insig2*), *Crnde* (denominado anteriormente *lincRNA-Irx5*), *Kantr* (denominado anteriormente *Spasm* y *lincRNA-Jarid1c*), *Pint* (denominado anteriormente *line-Pint* y *lincRNA-Mkln1*), *Trp53cor1* (denominado anteriormente *lincRNA-p21*), *lincppara* (denominado anteriormente *lincRNA-Ppara*), *Peril* (denominado anteriormente *lincRNA-Sox2*), *Tug1* (denominado anteriormente *lincRNA-Tug1*) y una combinación de los mismos.
- En un aspecto descrito en el presente documento, la célula o tejido que carece de una secuencia de ácido nucleico que codifica un ARNlinc funcional carece de un *Pint* funcional (denominado anteriormente *lincRNA-Mkln1*).
- En un aspecto descrito en el presente documento, se proporciona un constructo de ácido nucleico, que comprende un brazo de homología en dirección 5' y un brazo de homología en dirección 3', en donde los brazos de homología en dirección 5' y en dirección 3' se dirigen a un locus de ARNlinc, en donde el constructo es capaz de alterar la transcripción del ARNlinc, inactivando el ARNlinc o reemplazando el ARNlinc.
- En diversos aspectos descritos en el presente documento, los constructos de direccionamiento que se dirigen a un locus de ARNlinc comprenden una secuencia seleccionada de una secuencia Kozak, una secuencia que codifica una fracción detectable (por ejemplo, un indicador, por ejemplo, un indicador como se describe en el presente documento; con, por ejemplo, opcionalmente un promotor unido operativamente al mismo), una secuencia de ácido nucleico que codifica un marcador seleccionable (con, por ejemplo, un promotor unido operativamente al mismo) y una combinación del mismo. En un aspecto descrito en el presente documento, el indicador y/o marcador seleccionable está(n) flanqueado(s) con sitios de recombinación específica de sitio que están dispuestos para efectuar una delección de la secuencia de ácido nucleico que codifica el gen marcador seleccionable y/o la secuencia de ácido nucleico que codifica la fracción detectable. En un aspecto descrito en el presente documento, el constructo no comprende un promotor unido operativamente a la fracción detectable.
- En un aspecto descrito en el presente documento, se proporciona un método alterar un ARNlinc, que comprende insertar una secuencia de ácido nucleico dentro de un locus de ARNlinc, en donde la inserción altera la transcripción del ARNlinc, suprime una o más regiones codificantes de ARNlinc o suprime una secuencia codificante completa de un ARNlinc.
- En un aspecto descrito en el presente documento, se proporciona un método para hacer un animal no humano que comprende una alteración o inactivación de un ARNlinc, que comprende modificar un genoma de un animal no humano de manera que el animal no humano ya no expresa una versión funcional del ARNlinc. En un aspecto descrito en el presente documento, el método comprende una etapa de empleo de un vector de direccionamiento para alterar la transcripción del ARNlinc, para suprimir una o más regiones codificantes de ARNlinc o suprimir una secuencia codificante completa del ARNlinc en el genoma del animal no humano.

En un aspecto descrito en el presente documento, se proporciona un método para hacer un animal no humano que comprende una inactivación de una función de ARNlnc, que comprende modificar un genoma de una célula animal no humana pluripotente o totipotente para alterar la transcripción del ARNlnc, para suprimir uno o más exones que codifican unas regiones que codifican ARNlnc o suprimir una secuencia codificante completa del ARNlnc en el genoma de la célula; utilizar la célula como una célula donante e introducir la célula donante en un embrión hospedador para formar un complejo célula donante-embrión hospedador; y gestar el complejo célula donante-embrión hospedador en un animal no humano adecuado en condiciones adecuadas para la gestación, en donde tras la gestación se obtiene una progenie que comprende la inactivación de la función de ARNlnc. En un aspecto descrito en el presente documento, la progenie se cría hacia la homocigosis respecto a la inactivación de la función de ARNlnc.

En un aspecto descrito en el presente documento, se proporciona un método para hacer un animal no humano que comprende una inactivación de una función de ARNlnc, que comprende modificar un genoma de una célula somática o una célula germinal de un animal no humano para alterar la transcripción del ARNlnc, para suprimir una o más regiones codificantes de ARNlnc o para suprimir una secuencia codificante completa del ARNlnc en el genoma de la célula; emplear el genoma de la célula en un óvulo enucleado para formar un óvulo modificado; gestar el óvulo modificado en un animal no humano subrogado adecuado en condiciones adecuadas para la gestación; y obtener una progenie de animales no humanos que comprende la inactivación de ARNlnc. En un aspecto descrito en el presente documento, la progenie se cría hacia la homocigosis respecto a la inactivación de la función de ARNlnc.

En un aspecto descrito en el presente documento, se proporciona un método para hacer un animal no humano que comprende una modificación genética en al menos un locus de ARNlnc. Tal método comprende (a) poner en contacto una célula pluripotente con un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3'; en donde el constructo de direccionamiento sufre recombinación homóloga con el locus de ARNlnc en un genoma de la célula para formar una célula pluripotente modificada; (b) introducir la célula pluripotente modificada en un embrión hospedador; y (c) gestar el embrión de hospedador en una madre subrogada, en donde la madre subrogada produce progenie que comprende un locus de ARNlnc modificado, en donde dicha modificación genética da como resultado una pérdida de función de al menos un ARNlnc.

En un aspecto descrito en el presente documento, se proporciona un método para modificar un locus de ARNlnc en una célula pluripotente. Tal método comprende introducir en la célula pluripotente un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado con brazos de homología 5' y 3' que pueden sufrir recombinación homóloga con el locus de ARNlnc; e identificar una célula pluripotente modificada que comprende una modificación genética dirigida al locus de ARNlnc, en donde la modificación genética da como resultado la pérdida de función de la función de ARNlnc. En un aspecto descrito en el presente documento, la célula pluripotente es una célula pluripotente inducida humana (iPS).

En diversos aspectos descritos en el presente documento, las células modificadas incluyen, por ejemplo, células pluripotentes, células pluripotenciales inducidas, células madre, células madre embrionarias, etc. En un aspecto específico descrito en el presente documento, la célula es una célula madre embrionaria (ES). En un aspecto específico descrito en el presente documento, la célula ES es una célula ES de ratón o de rata.

En diversos aspectos descritos en el presente documento, los animales no humanos incluyen, por ejemplo, especies de animales ovinas, bovinas, porcinas y murinas. En un aspecto específico descrito en el presente documento, el animal es de una especie murina, por ejemplo, un ratón o una rata.

Se incluyen en la divulgación otros aspectos adicionales, como se apreciará por los expertos en la técnica tras la lectura de la presente divulgación.

Breve descripción de las figuras

El archivo de patente o solicitud contiene al menos un dibujo realizado a color. La Oficina proporcionará copias de esta patente o publicación de solicitud de patente con dibujos a color previa solicitud y pago de la tarifa necesaria.

La **Figura 1** ilustra una estrategia para la alteración dirigida del locus del gen *Fendrr*. Se muestra un mapa parcial del locus *Fendrr* de ratón de tipo silvestre, que incluye los exones 1 - 6. Tras la recombinación homóloga, el LTVEC de direccionamiento reemplazó un total de 19,2 kb de la secuencia genómica *Fendrr* con el casete LacZ-resistencia a la neomicina que introduce una secuencia genómica Kozak. Las cajas vacías indican exones no codificantes. Las cajas rojas y verdes en el locus genómico *Fendrr* y en el casete LacZ-resistencia a la neomicina están las secuencias homólogas utilizadas para el direccionamiento.

La **Figura 2** ilustra la expresión génica del indicador LacZ espacial y temporal en embriones de ratón de ARNlnc dirigido en la etapa media de gestación **A**. Los embriones e12,5 heterocigotos fijados y teñidos para β-galactosidasa mostraron una amplia gama de expresión del gen indicador LacZ introducido en el cerebro en desarrollo y región craneofacial (por ejemplo, *Pantr1* y *Pantr2*, *Celrr* y *Haglr*, véase también la Fig. 9), tubo neural (*Pantr2*, *Halr1* y *lincppara*), aorta dorsal (*Celrr*), corazón (*Celrr*, *Haglr*, véase también la Fig. 9), pulmones (*Fendrr*), brote de las extremidades (*HOTTIP*, *Hoxallos* y *Mannr*), intestino delgado (*HOTTIP*, *Hoxallos* y *Fendrr*), región

posterior y la cola (*HOTAIR*, *HOTTIP* y *Hoxallos*). Análisis similares mostraron un patrón de expresión de lacZ extendido en *Tug1*, si bien la expresión de otros genes indicadores pudo quedar restringida a la epidermis (*Eldr*), brotes mamarios (*Lincenc1*, véase también la Fig. 9) o placoda de bigotes (*Trp53cor1*, véase también la Fig. 9). Los ejemplos que se muestran son representativos de al menos cinco embriones de genotipos confirmados por proyecto de inactivación de ARNlinc. **B**, Los patrones de expresión en embriones heterocigóticos F1 de ARNlinc seleccionados (*HOTTIP*, *Hoxallos* y *Celrr*) de las etapas indicadas (e9,5-e12,5) mostraron que la expresión comenzó de forma temprana en un sitio restringido y a continuación se extendió más allá del sitio inicial en etapas posteriores. La expresión de *Celrr* estuvo circunscrita al cerebro en la e9,5 y progresó en la médula espinal en la e12,5. La expresión de *Hoxallos* comenzó en el brote de la cola en desarrollo y progresa en la región caudal completa del embrión, extremidad trasera y extremidad delantera en la e12,5. La expresión de *HOTTIP* también comenzó en el brote de la cola en desarrollo y a continuación se observó en los autópodos distales en desarrollo de la extremidad anterior y de la extremidad trasera en la e11,5 y e12,5. Los ejemplos mostrados son representativos de al menos 5, embriones de genotipo confirmado por proyecto de ARNlinc.

La **Figura 3** ilustra la expresión génica del indicador LacZ (azul) en cerebros de heterocigóticos de generación F0 de ARNlinc de 6-8 semanas de edad. **A**, *Celrr*, expresión amplia en materia gris con la excepción del cerebelo lateral y protuberancias ventrales; **B**, *Crnndc*, expresión en los colículos (vista dorsal, flecha); **C**, *Pantr1*, expresado en el neocórtex, bulbo olfatorio, cerebro anterior basal e hipotálamo; **D**, *Pantr2*, expresado en el neocórtex, bulbo olfatorio, cerebelo, hipotálamo y cerebro anterior basal; **E**, *Lincenc1*, expresado en neocórtex, partes del cerebelo e hipotálamo medial con fuerte patrón de expresión especialmente en la proyección olfatoria y áreas de proyección olfatoria del córtex temporal (vista ventral, flecha roja); **F**, *Pint*, expresado ubicuamente en la materia gris con expresión especialmente intensa en el hipotálamo; **G**, *lincppara*, expresado ampliamente en la materia gris con expresión especialmente densa en el hipotálamo; **H**, *Peril*, expresión en la línea media del hipotálamo (vista ventral, punta de flecha); **I**, *Kantr*, alguna expresión potencialmente en las capas cerebelares profundas (vista dorsal, estrella); y **J**, *Tug1*, expresión en la materia gris de la médula espinal y ligera expresión de materia gris en la mayoría de estructuras excepto para el neocórtex. n = 2, ratones macho de genotipo confirmado para el proyecto de inactivación de ARNlinc.

La **Figura 4** ilustra que *Pint* muestra una expresión incrementada desde el día 3 posnatal a las 8 semanas de edad. La expresión génica del indicador LacZ (azul) a los 3 días, 3 semanas y 8 semanas en heterocigotos F0 muestra que *Pint* se expresa de forma incrementada con la edad. **A**, A los 3 días, la tinción de β -galactosidasa solo se observa en partes del cerebro, tendones y ligamentos de la extremidad trasera y algunos bronquiolos en el pulmón (flecha). **B**, A las 3 semanas, hay una tinción incrementada en el cerebro, pata trasera, atrio del corazón, pulmón e hígado. **C**, A las 8 semanas de edad, el cerebro completo, músculo esquelético de la pata trasera y tórax, atrio y miocardio, pulmón y tejido hepático, mostraron una tinción de β -galactosidasa fuerte representativa de la expresión de *Pint* incrementada. Los ejemplos mostrados son representativos de n>4 ratones por grupo.

La **Figura 5** ilustra el fenotipo asociado con envejecimiento prematuro en ratones nuligénicos (en inglés, knockout) para *Pint*. (A) Los machos de ratones *Pint*^{-/-} y *Pint*^{+/-} muestran una tasa de crecimiento significativamente más pequeña que sus compañeros de camada de tipo silvestre (WT; del inglés, wild type) y comienzan a mostrar una pérdida de peso significativa cerca de los 6 meses de edad. Los datos se representan como la media +/- ETM, n > 9 ratones para cada grupo. El nivel de significación se calculó mediante un ANOVA de una vía (*, P < 0,05; **, P < 0,005; ***, P < 0,001). (B) Análisis de Kaplan-Meier de ratones heterocigóticos con homocigóticos y WT. Los ratones macho *Pint*^{-/-} mostraron una reducción significativa en la supervivencia en comparación con *Pint*^{+/-} y compañeros de camada de tipo silvestre. Los datos se representaron como porcentaje de supervivencia a lo largo de 1 año de observación. (C) Secciones de piel ventral y dorsal en ratones *Pint*^{-/-} en comparación con *Pint*^{+/-} y compañeros de camada WT. (D, E, F y G) Evaluación por MicroCT de la composición corporal a las 12, 26 y 52 semanas de edad. (D, E) Los ratones macho *Pint*^{-/-} y *Pint*^{+/-} muestran una reducción significativa en grasa corporal ya a las 26 semanas de edad. Los ratones hembra *Pint*^{-/-} han reducido grasa corporal a una edad mayor considerablemente a las 52 semanas de edad (***, P < 0,001, ANOVA de una vía). (F, G) Se observó una reducción significativa en la densidad mineral ósea (DMO; en inglés, BMD) tanto en machos como hembras *Pint*^{-/-} en comparación con sus compañeros de camada *Pint*^{+/-} y WT (*, P < 0,05; ***, P < 0,001, ANOVA de una vía). (H) Las imágenes de MicroCT muestran una cifolordosis pronunciada (curvatura de la columna vertebral) observada en ratones machos y hembras *Pint*^{-/-} mayor en comparación con compañeros de camada WT. (I) Aproximadamente el 70 % de los ratones machos y hembras *Pint*^{-/-} tienen cifolordosis a las 12 semanas de edad, en comparación con el 0-20 % de compañeros de camada *Pint*^{+/-} y WT. A las 26 semanas de edad, la proporción de ratones *Pint*^{-/-} con cifolordosis se incrementó en casi el 90 % y apareció en aproximadamente el 60 % de ratones *Pint*^{-/-}, en comparación con menos del 20 % de compañeros de camada WT. n > 9 ratones por grupo para todas las observaciones informadas.

La **Figura 6** ilustra que los ratones nuligénicos para *Fendrr* mostraron una morfología de pulmón anormal en la e13,5. **A**. La expresión génica del indicador LacZ en la e12,5 en embriones KO (del inglés, knockout) para *Fendrr* muestra una expresión positiva en la región frontonasal (FN) de la cara, la región aorta gónada mesonefros (AGM) y el tracto respiratorio incluyendo los pulmones (L; del inglés, lung) y la tráquea (T). **B**. La disección de los pulmones en la e13,5 reveló un fenotipo anormal, desestructurado, globular en los lóbulos del KO en comparación con el Het.

La **Figura 7** ilustra la transformación homeótica observada en la 4ª vértebra caudal de ratones KO para *HOTAIR*. **A.** La visualización de la región sacra y caudal del esqueleto de ratón mediante μ CT revela una transformación homeótica de la 4ª vértebra caudal en una estructura similar a la de la 3ª vértebra caudal en ratones KO para *HOTAIR*. **B.** La comparación de la 4ª vértebra caudal dorsal, lateral y ventral de KO WT y *HOTAIR* revela una anomalía estructural en KO indicativa de una transformación homeótica.

La **Figura 8** ilustra que los ratones KO para *HOTTIP* mostraron una postura de las patas traseras anormal, una fuerza de prensión reducida en las patas delanteras y traseras y un fenotipo de desgaste muscular. **A.** Los ratones KO para *HOTTIP* mostraron una postura inusual de "juntado" de las patas traseras cuando se suspendieron de la cola. WT, tipo silvestre; KO, nuligénico. **B.** La prueba de resistencia a la jaula reveló que los ratones KO para *HOTTIP* tienen una capacidad reducida para permanecer suspendidos de la parte superior de una jaula de alambre invertida. n=5 ratones para cada grupo. **C.** Se tomaron los músculos TA (tibial anterior) derecho e izquierdo, GA (gastrocnemio) y Quad (cuádriceps) de ratones WT, Het y KO y se pesaron. Los pesos musculares se normalizaron por el peso corporal y se calcularon para incluir los pesos de los músculos derecho/izquierdo. Los datos son la media \pm ETM, n= 6 ratones para cada grupo. Se observó una disminución significativa en el peso muscular solo en el GA del animal KO para *HOTTIP* tanto en machos como en hembras (datos de machos no mostrados). Los asteriscos indican una diferencia significativa en los pesos musculares de GA de KO en comparación con todos los otros grupos de control (P < 0.01). **D.** Comparación de las cifras de fibras musculares de GA en WT, Het y KO. Se observó una reducción significativa del recuento de fibras en los KO. El nivel de significación se calculó utilizando un ANOVA de una vía (P < 0,0001). **E.** Comparación del área de fibras musculares media en sección transversal. Las secciones transversales tomadas del músculo GA se tiñeron con un anticuerpo contra laminina (Sigma) y se midieron. No hay una diferencia de tamaño considerable entre los músculos esqueléticos de KO y control. n = 6 ratones por grupo para todos los análisis musculares.

La **Figura 9** ilustra la tinción precisa en áreas pequeñas para cuatro inactivaciones de ARNinc que muestran patrones de tinción altamente específicos: Perfiles de expresión de *lacZ* a media gestación específicos para *Peril*, *Ptgs2os2*, *Trp53cor1* y *Lincenc1* (A) El perfilado del indicador *LacZ* para *Peril* muestra un perfil de expresión neuronal específico así como una fuerte expresión en el corazón y región posterior de la cola. (B) La expresión del indicador *Ptgs2os2 lacZ* se restringe a la base de las patas anteriores y traseras en desarrollo. (C) La expresión del indicador *Trp53cor1 lacZ* es específica de la placoda de bigotes en el proceso nasal en desarrollo. Los embriones e12,5 recogidos de la misma camada recogieron la progresión del desarrollo de placoda de bigotes durante un periodo de tiempo corto. (D) Se extirparon las patas delanteras y traseras en los embriones *Lincenc1*^{+/−} para revelar la expresión de brotes mamarios (puntas de flecha). Vista ventral del embrión *Lincenc1*^{+/−} e12,5: se detecta expresión de *lacZ* en cinco pares de brotes mamarios.

La **Figura 10** ilustra un fenotipo caracterizado por la pérdida de hueso del talón en nuligénicos para *HOTTIP*. Malformaciones esqueléticas observadas en ratones mutantes *Hottip*. Además de un fenotipo de músculo esquelético en la pata delantera, los ratones *Hottip*^{−/−} también mostraron una anomalía de hueso esquelético observada mediante microCT 3D. Los ratones *Hottip*^{−/−} tanto machos como hembras (C y F) tienen calcáneos acortados (flechas) en comparación con WT (A y D) y los controles de sus compañeros de camada *Hottip*^{+/−} (B y E).

La **Figura 11** representa una tabla (Tabla 2) de la expresión de indicadores en embriones y tejido de adultos para las inactivaciones de ARNinc del estudio.

Descripción detallada de la invención

Glosario

La expresión "célula madre embrionaria" o "célula ES" incluye una célula totipotente o pluripotente derivada de un embrión que es capaz de contribuir a cualquier tejido del embrión en desarrollo tras la introducción en un embrión. La expresión "célula pluripotente" incluye una célula indiferenciada que posee la capacidad de desarrollarse en más de un tipo celular diferenciado.

La expresión "vector de direccionamiento grande" o "LTVEC" (del inglés, large targeting vectors for eukaryotic cells) incluye vectores de direccionamiento grandes para células eucariotas que se derivan de fragmentos de ADN genómico clonados mayores que aquellos usados normalmente por otros enfoques previstos para realizar direccionamiento de genes homólogos en células eucariotas. Entre los ejemplos de LTVEC, se incluyen, pero no se limitan a, cromosoma homólogo bacteriano (BAC; del inglés, bacterial homologous chromosome) y cromosoma artificial de levaduras (YAC; del inglés, yeast artificial chromosome).

La expresión "sitio de recombinación" incluye una secuencia de nucleótidos que se reconoce por una recombinasa específica de sitio y que puede servir como un sustrato para un evento de recombinación.

La expresión "recombinasa específica de sitio" incluye un grupo de enzimas que pueden facilitar la recombinación entre "sitios de recombinación". Entre los ejemplos de "recombinasa específica de sitio" se incluyen, pero no se limitan

a, las recombinasas Cre, Flp y Dre.

La expresión "línea germinal" en referencia a una secuencia de ácido nucleico, incluye una secuencia de ácido nucleico que se puede pasar a la progenie.

5 La expresión "unido operativamente" significa que se unen componentes para funcionar juntos en su manera prevista. En un caso, una secuencia de ácido nucleico que codifica una proteína puede unirse operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) con el fin de conservar la regulación de la transcripción adecuada.

10 La expresión "ARN no codificante largo" o "ARNlnc" como se utiliza en el presente documento incluye un transcrito no codificante de proteínas de más de 200 nucleótidos de largo.

15 La expresión "ARN no codificante intergénico largo" o "ARNlinc" como se utiliza en el presente documento incluye un subgrupo de ARNlnc. Tal como se usa en el presente documento, los ARNlinc no solapan con exones de regiones codificantes de proteínas del genoma.

20 El término "locus" se define como un segmento de ADN dentro del ADN genómico. Por ejemplo, un locus de ARNlnc es un segmento de ADN dentro del ADN genómico que codifica un ARNlnc.

1. Composiciones que comprenden modificación genética de al menos un locus de ARNlnc

25 En el presente documento se describen animales no humanos, células, tejidos y embriones que comprenden pérdida de función de ARNlnc, que incluyen, pero sin limitarse a las alteraciones o inactivaciones de uno o más ARNlnc. Se describen en el presente documento métodos y composiciones para manipular la expresión de ARNlnc. Las composiciones de direccionamiento dirigidas a modificar o inactivar ARNlnc también se describen en el presente documento. En el presente documento también se proporcionan animales no humanos, células y tejidos que muestran un fenotipo asociado con la no función de uno o más ARNlnc. Aunque la siguiente descripción se hace en referencia a una investigación de ciertos ARNlnc en particular, los métodos y composiciones pueden practicarse con cualquier ARNlnc.

35 En el presente documento se describen animales no humanos, células, tejidos y embriones que comprenden una modificación genética dirigida en al menos un locus de ARN no codificante largo (ARNlnc). En tales casos, el locus de ARNlnc modificado comprende una mutación de pérdida de función en una secuencia de ácido nucleico que codifica el ARNlnc. En el presente documento, también se describen células, tejidos y embriones derivados de animales no humanos que comprenden una mutación de pérdida de función de al menos un ARNlnc.

40 La expresión, "pérdida de función" en lo referido a un ARNlnc puede incluir cualquier modificación en un locus de ARNlnc que da como resultado una disminución o ausencia de la expresión del ARNlnc y/o una disminución o ausencia de actividad/función del ARNlnc. El nivel de expresión de un ARNlnc se puede medir directamente, por ejemplo, evaluando el nivel del ARNlnc en la célula u organismo.

45 En general, el nivel de expresión y/o actividad del ARNlnc disminuye(n) si el nivel de expresión de ARNlnc y/o el nivel de actividad del ARNlnc es(son) estadísticamente menor(es) ($p \leq 0,05$) que el nivel de ARNlnc en una célula u organismo de control apropiado que no ha sido modificado genéticamente o mutagenizado para inhibir la expresión y/o actividad del ARNlnc. En aspectos específicos descritos en el presente documento, la concentración y/o actividad del ARNlnc disminuye(n) en al menos un 1 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 99 % o más respecto a una célula u organismo de control que no ha sido modificado para tener el nivel y/o actividad de ARNlnc disminuido(s).

50 En otros casos, se seleccionan células u organismos que tienen la modificación genética dirigida que reduce el nivel de expresión y/o la actividad del ARNlnc, utilizando métodos que incluyen, pero no se limitan a, análisis de transferencia de Southern, secuenciación de ADN, análisis de PCR o análisis fenotípico. A continuación, se emplean tales células u organismos en los diversos métodos y composiciones descritas en el presente documento.

55 Una "célula sujeto" u "organismo sujeto" es uno en el que una alteración genética, tal modificación genética desvelada en el presente documento se ha efectuado o es una célula/organismo que desciende de una célula/organismo así alterado y el cual comprende la alteración. Un "control" o "célula de control" u "organismo de control" proporciona un punto de referencia para medir cambios en el fenotipo de la célula u organismo sujeto. En un aspecto descrito en el presente documento, una célula/organismo de control está tan emparejada/o estrechamente como es posible con la célula/organismo con la modificación genética en el ARNlnc, excepto en que carece de la modificación o mutación genética, dando como resultado la expresión y/o actividad reducida(s) (por ejemplo, las células respectivas pueden originarse de la misma línea celular). En otros casos, la célula/organismo de control puede comprender, por ejemplo: 60 (a) una célula/organismo de tipo silvestre, es decir, del mismo genotipo que el material de partida para la alteración genética que dio como resultado la célula/organismo sujeto; (b) una célula/organismo del mismo genotipo que el material de partida pero que se ha modificado genéticamente con un constructo nulo (es decir, con un constructo que 65

no tiene efecto conocido sobre la característica de interés, tal como un constructo que comprende un gen marcador); (c) una célula/organismo que es una progenie no modificada genéticamente de una célula/organismo sujeto (es decir, la célula de control y la célula sujeto se originan de la misma línea celular); (d) una célula/organismo idéntica/o genéticamente a la célula/organismo pero que no está expuesta/o a condiciones o estímulos que pudieran inducir la expresión del gen de interés; o (e) la célula/organismo sujeto en sí misma/o, en condiciones en las que la modificación genética no da como resultado una alteración en la expresión del polinucleótido de interés.

El término "animal", en referencia a animales, células, tejidos o embriones, incluye mamíferos, peces y aves. Los mamíferos incluyen, por ejemplo, seres humanos, primates no humanos, mono, simio, gato, perro, caballo, toro, ciervo, bisonte, oveja, roedores (por ejemplo, ratones, ratas, hámsteres y cobayas), ganado (por ejemplo, especies bovinas, por ejemplo, vacas, buey, etc.; especies ovinas, por ejemplo, oveja, cabras, etc.; y especies porcinas, por ejemplo, cerdos y jabalíes). Las aves incluyen, por ejemplo, pollos, pavos, avestruz, gansos, patos, etc. Se incluyen también animales domesticados y agrícolas. La expresión "animal no humano", en referencia a animales, células, tejidos o embriones, excluye a seres humanos.

En un aspecto descrito en el presente documento, el animal es un animal no humano. En otro aspecto descrito en el presente documento, el animal no humano es un mamífero. En otro aspecto descrito en el presente documento, el mamífero es un roedor. En un aspecto adicional descrito en el presente documento, el roedor es un ratón, una rata o un hámster.

Las modificaciones genéticas como se describen en el presente documento pueden incluir una o más deleciones de un locus de ARNlnc de interés, adiciones a un locus de ARNlnc de interés, reemplazo de un locus de ARNlnc de interés y/o cualquier combinación de los mismos. El locus de interés puede comprender regiones codificantes o regiones reguladoras no codificantes.

Las modificaciones genéticas proporcionadas en el presente documento se dirigen a un locus de ARNlnc de interés. Una pérdida de función de un ARNlnc puede resultar de una modificación genética dirigida en el gen de ARNlnc (es decir, una modificación genética en una región reguladora, la región codificante, exones y/o intrones, etc.). Tales modificaciones dirigidas incluyen, pero no se limitan a, adiciones de uno o más nucleótidos, deleciones de uno o más nucleótidos, sustituciones de uno o más nucleótidos, una alteración del locus de ARNlnc, una inactivación del locus de ARNlnc o una porción del mismo, una activación del locus de ARNlnc o una porción del mismo, un reemplazo de una secuencia de ácido nucleico de ARNlnc endógena o una porción de la misma con una secuencia de ácido nucleico heteróloga o una combinación de la misma. En aspectos específicos descritos en el presente documento, al menos 1, 2, 3, 4, 5, 7, 8, 9, 10, 50, 100, 400 o más nucleótidos se cambian para formar la modificación genómica dirigida.

En un aspecto descrito en el presente documento, la mutación de pérdida de función se caracteriza por una alteración o una inactivación de al menos una función de ARNlnc.

El locus de ARNlnc se puede modificar genéticamente en cualquier región del locus, de manera que la modificación da como resultado una pérdida de función del ARNlnc. En un aspecto descrito en el presente documento, la modificación del locus de ARNlnc comprende una deleción de la región codificante de ARNlnc completa o una porción de la misma. En un aspecto descrito en el presente documento, el locus de ARNlnc modificado comprende una deleción de uno o más exones que codifican el ARNlnc o una porción del mismo. En otro aspecto descrito en el presente documento, la deleción comprende una deleción de uno o más exones dentro del locus de ARNlnc que comienza en un primer exón del locus de ARNlnc. En otros aspectos descritos en el presente documento, la deleción comprende una deleción de uno o más exones dentro del locus de ARNlnc que comienza en un segundo exón del locus de ARNlnc.

En algunos casos, se reemplaza el locus de ARNlnc o una porción del mismo con un inserto de ácido nucleico. En tales casos, el reemplazo puede ser un reemplazo de la región codificante de ARN completa del locus de ARNlnc o una porción de la misma con el inserto de ácido nucleico, un reemplazo de uno o más exones del locus de ARNlnc con el inserto de ácido nucleico, un reemplazo de uno o más exones dentro del locus de ARNlnc comenzando en el primer exón del locus de ARNlnc con el inserto de ácido nucleico o un reemplazo de uno o más exones dentro del locus de ARNlnc que comienza en el segundo exón con el inserto de ácido nucleico.

En algunos casos, el inserto de ácido nucleico se posiciona en el locus de ARNlnc de manera que está en unión operativa con un promotor de ARNlnc endógeno de manera que el promotor de ARNlnc endógeno impulsa la expresión del inserto de ácido nucleico. En tales casos, la expresión de la secuencia de ácidos nucleicos sigue un patrón de expresión del ARNlnc.

En un aspecto descrito en el presente documento, el locus de ARNlnc o porción del mismo se reemplaza con un inserto de ácido nucleico que comprende una primera secuencia de ácido nucleico que codifica un indicador. Por ejemplo, en el caso donde el inserto de ácido nucleico comprende un gen indicador y se sitúa en el locus de ARNlnc en unión operativa para el promotor de ARNlnc, la expresión del gen indicador está impulsada por el promotor de ARNlnc endógeno. Como alternativa, el inserto de ácido nucleico no se inserta en unión operativa con el promotor de ARNlnc endógeno. En tales casos, el inserto de ácido nucleico puede comprender un promotor. En un aspecto descrito en el

presente documento, el inserto de ácido nucleico comprende un gen indicador unido operativamente a un promotor que impulsa la expresión del gen indicador.

5 En un aspecto descrito en el presente documento, el locus de ARNlnc o porción del mismo se reemplaza con un inserto de ácido nucleico que comprende una segunda secuencia de ácido nucleico que codifica un marcador seleccionable. En tales casos, la segunda secuencia de nucleótidos está ligada operativamente a un promotor que impulsa la expresión del marcador seleccionable.

10 En otro aspecto descrito en el presente documento, el locus de ARNlnc o porción del mismo se reemplaza con un inserto de ácido nucleico que comprende un gen indicador y un gen marcador seleccionable. En tales casos, el gen indicador y/o el gen marcador seleccionable puede(n) o no estar unido(s) operativamente a un promotor.

Se proporcionan en cualquier otra parte en el presente documento, varios promotores que se pueden emplear en los métodos y composiciones.

15 Tales modificaciones genéticas (incluyendo aquellas que dan como resultado una disminución o una modulación en la expresión y/o actividad del ARNlnc diana) son capaces también de transmitirse a través de la línea germinal. En aspectos específicos descritos en el presente documento, las modificaciones genéticas dan como resultado una inactivación del locus diana deseado. Tales animales no humanos, por ejemplo, encuentran uso en diversos sistemas experimentales como se analiza en otra parte en el presente documento.

20 Por ejemplo, las inactivaciones de ARNlnc ofrecen un modelo animal para estudiar la función de ARNlnc, el papel del ARNlnc en el desarrollo y el papel del ARNlnc en diversas rutas y enfermedades, incluyendo, pero sin limitarse a, envejecimiento, incluyendo el envejecimiento prematuro, desarrollo cerebral, desarrollo embrionario, desarrollo pulmonar, desarrollo esquelético, desarrollo muscular, cáncer o regulación de la transcripción.

25 Se pueden utilizar diversos métodos para generar la modificación genética dirigida y están descritos en otra parte en el presente documento.

30 A. ARNlnc

Los animales no humanos, células, tejidos y embriones empleados en los métodos y composiciones proporcionados en el presente documento tienen una modificación genética que da como resultado la pérdida de función de al menos un ARNlnc. Los ARNlnc son ARN no codificantes largos de más de 200 nucleótidos. Un subgrupo de ARNlnc, ARN no codificante intergénico largo (ARNlinc) es intergénico y no solapa con regiones codificantes de proteínas.

40 Cualquier locus de ARNlnc se puede modificar en los métodos y composiciones proporcionados en el presente documento. En un aspecto descrito en el presente documento, el animal no humano modificado genéticamente, célula, tejido o embrión comprende una modificación genética en un ARNlnc. En otro aspecto descrito en el presente documento, el ARNlnc es un ARNlinc.

45 Los ejemplos no limitantes de un ARNlnc comprenden *HOTAIR*, *HOTTIP*, *Hoxallos* (denominado anteriormente *HoxA11as*), *Pantr1* (denominado anteriormente *lincRNA-Brn1-a*), *Pantr2* (denominado anteriormente *lincRNA-Brn1-b*), *Ptgs2os2* (denominado anteriormente *lincRNA-Cox2*), *Eldr* (denominado anteriormente *FabI* y *lincRNA-Egfr*), *lincenc1* (denominado anteriormente *lincRNA-Enc1*), *Mannr* (denominado anteriormente *lincRNA-Evil*), *Fendrr* (denominado anteriormente *lincRNA-Foxf1*), *Halr1* (denominado anteriormente *Haunt* y *lincRNA-HoxA1*), *Haglr* (denominado anteriormente *Mdgt* y *lincRNA-HoxD3*), *Celrr* (denominado anteriormente *Celr* y *lincRNA-Insig2*), *Cmde* (denominado anteriormente *lincRNA-Irx5*), *Kantr* (denominado anteriormente *Spasm* y *lincRNA-Jarid1 c*), *Pint* (denominado anteriormente *line-Pint* y *lincRNA-Mklnl*), *Trp53cor1* (denominado anteriormente *lincRNA-p21*), *lincppara* (denominado anteriormente *lincRNA-Ppara*), *Peril* (denominado anteriormente *lincRNA-Sox2*), *Tug1* (denominado anteriormente *lincRNA-Tug1*) o una combinación de los mismos.

55 Ahora es evidente que los genes que codifican proteínas no son todos los que hay en el genoma (Mattick, J.S. (2009), *PLoS Genet* 5:e1000459). Los estudios a gran escala de expresión del genoma completo en células de mamífero han revelado que aproximadamente tres cuartas partes del genoma es capaz de expresarse como ARN (Carninci, P., *et al.* (2005), *Science* 309:1559-1563; Djebali, S., *et al.* (2012), *Nature* 489:101-108; Kapranov, P., *et al.* (2007), *Science* 316:1484-1488) y que la mayoría de los transcritos no codifican para proteínas. Entre los transcritos no codificantes existe una clase diversa conocida como los ARNlnc no codificantes (ARNlnc). Representando aproximadamente 15.000 transcritos de los cerca de 10.000 loci genómicos en células humanas (Derrien, T., *et al.* (2012), *Genome Res* 22:1775-1789.), los ARNlnc y una subclase conocida como ARN no codificante intergénico grande (ARNlinc) (Guttman, M., *et al.* (2009), *Nature* 458:223-227; Khalil *et al.* (2009)) se asemejan a ARNm codificante de proteínas en estructura, síntesis y el carácter cromatínico de sus genes. Si esta similitud estructural se extiende o no hacia una diversidad funcional que empareje proteínas permanece como una cuestión abierta.

65 Los estudios funcionales sobre ARNlnc individuales han identificado papeles en la inactivación cromosómica de X (Marahrens, Y., *et al.* (1997), *Genes Dev* 11:156-166), impronta (Leighton, P.A., *et al.* (1995), *Nature* 375:34-39;

Mohammad, F., *et al.* (2010), *Development* 137:2493-2499; Sleutels, F., *et al.* (2002), *Nature* 415:810-813; Takahashi, N., *et al.* (2009), *Hum Mol Genet* 18:1879-1888), diferenciación retinal (Young, T.L., *et al.* (2005), *Curr Biol* 15:501-512) y desarrollo del corazón y de la pared corporal (Grote, P., *et al.* (2013), *Dev Cell* 24:206-214). Los estudios sobre el ARNlinc *HOTAIR* primero revelaron que los ARNlinc podrían regular la expresión génica en sitios lejos de sus propios sitios de transcripción guiando complejos modificadores de cromatina (complejo 2 represor polycomb en el caso de *HOTAIR*) hacia loci genómicos específicos (Rinn, J.L., *et al.* (2007), *Cell* 129:1311-1323). Se han encontrado mecanismos de acción similares para el ARNlinc Xist en la inactivación del cromosoma X (Zhao, J., *et al.* (2008), *Science* 322:750-756) y para los ARNlinc AIR y Kcnq1ot1 en la impronta. Estos hallazgos sugieren un papel más amplio para los ARNlinc en la regulación de la expresión génica, que se ha sustentado en análisis de patrones de expresión correlacionados para ARNlinc y genes que codifican proteínas que señalan a la participación de ARNlinc en una matriz de amplia gama de procesos celulares y fisiología del sistema de órganos (Guttman *et al.* (2009)). Muchos de los estudios recientes sobre ARNlinc han empleado estrategias genómicas globales que han establecido una visión general del papel de los ARNlinc como una clase. Para responder a las cuestiones de si las acciones de los ARNlinc sobre la expresión génica de proteínas son amplias, sutiles y tamponadoras o específicas, directas y determinantes, requiere la investigación de sus papeles individuales en animales vivos.

En el presente documento se proporcionan, en la siguiente descripción, ejemplos no limitativos de modificaciones genéticas que dan como resultado la inactivación de varios ARNlinc en un modelo de ratones nuligénicos. Se realizó una investigación de la expresión génica y de fenotipos en ratones nuligénicos para 20 genes de ARNlinc, que incluyeron el perfilado de *LacZ* que mostró diversos patrones espaciotemporales de expresión específica de tejidos; se revelaron dos líneas de inactivación mostraron letalidad perinatal; y se revelaron otros fenotipos que incluyen un fenotipo asociado a envejecimiento prematuro y defectos en pulmón, esqueleto, cerebro y músculo.

Para comenzar a examinar las funciones de los ARNlinc en un escenario de animales vivos, se crearon líneas de ratones nuligénicos para veinte genes de ARNlinc. Cada alelo mutante llevó un indicador *lacZ* cuyo perfil de expresión reveló un amplio espectro de patrones de expresión espaciotemporales y específicos de tejidos en adultos y embriones. Entre las 18 líneas de inactivación homocigotas, seis (aproximadamente un 33 %) mostraron fenotipos mutantes discernibles, dos (11 %) de los cuales fueron letales perinatales e incluyeron un fenotipo asociado al envejecimiento prematuro, morfología aberrante en el cerebro, pulmones, esqueleto y músculo y cambios globales en los patrones de expresión génica, que juntos señalan a diversos papeles para esta nueva clase de ARN funcionales en el desarrollo embrionario y en la fisiología de una matriz amplia de tejidos y órganos.

B. Modificación genética de un locus de ARNlinc

En el presente documento se proporcionan métodos y composiciones para la modificación genética de al menos un locus de ARNlinc en un animal no humano, célula, tejido o embrión.

La modificación genética del locus de ARNlinc de interés puede ser cualquier modificación del locus como se describe en detalle en otra parte en este documento (es decir, delección, inserción, reemplazo, etc.). En tales casos, la modificación genética da como resultado la pérdida de función del ARNlinc. En un aspecto descrito en el presente documento, la modificación genética comprende una alteración o una inactivación de al menos un ARNlinc.

i. Diseño y construcción del alelo de inactivación

El diseño y la construcción de un alelo de ARNlinc modificado, tal como una inactivación, es complicado debido a varios problemas técnicos. Por ejemplo, hay una carencia general de relaciones estructura-función para los ARNlinc y un locus de ARNlinc no tiene un marco de lectura abierto. Por tanto, las mismas estrategias que guiarían el diseño de un alelo para modificar una secuencia codificante de proteínas, tal como una inactivación, pueden no ser aplicables al ARNlinc. Además, los límites de los genes ARNlinc no están bien definidos, lo cual complica adicionalmente el diseño de un alelo de ARNlinc modificado, tal como una inactivación. En el presente documento, más adelante, se describen en detalle ejemplos no limitativos de estas dificultades técnicas y las estrategias utilizadas en el presente documento para superar satisfactoriamente estos obstáculos en el diseño de inactivación de ARNlinc.

En un ejemplo, se aplicaron los métodos y composiciones proporcionados en el presente documento a ratón como modelo de sistema. Sin embargo, aunque la siguiente descripción hace referencia a ratones, se puede utilizar cualquier animal no humano, célula, tejido o embrión en los métodos y composiciones descritos en el presente documento.

Desde su invención hace casi veinticinco años, el método para crear ratones modificados genéticamente, denominados nuligénicos (en inglés, knockout), ha establecido al ratón como el sistema más importante para el estudio de la función génica en mamíferos (Capecchi, M.R. (2001), *Nat Med* 7:1086-1090; Evans, M.J. (2001), *Nat Med* 7:1081-1083; Smithies, O. (2001), *Nat Med* 7:1083-1086). Con pocas excepciones, la aplicación de tecnología de ratones nuligénicos tanto en estudios genéticos individuales como en proyectos internacionales a gran escala (Bradley, A., *et al.* (2012), *Mamm Genome* 23:580-586) se ha centrado en los genes codificantes de proteínas, pero los esfuerzos recientes para crear recursos de ratones nuligénicos globales para microARN (Prosser, H. M., *et al.* (2011), *Nat Biotechnol* 29:840-845) (mcmanuslab.ucsf.edu/microrna_knockout) demuestran el valor de la aplicación de la tecnología de ARN no codificantes.

La aplicación de tecnología de ratones nuligénicos a ARNlnc, sin embargo, presenta algunas cuestiones técnicas y dificultades. La mayoría de las proteínas tienen elementos o dominios que son conocidos por ser o al menos se prevén importantes para una función. La delección de secuencias codificantes para estas partes esenciales a menudo es suficiente para crear un alelo nulo. Asimismo, se pueden diseñar alelos condicionales que aislen el exón o exones críticos para una delección posterior mediante la acción de una recombinasa específica de tejidos. Debido a que no se han establecido todavía relaciones estructura-función para todos menos para algunos ARNlnc y a que no hay marco de lectura abierto como una guía, las estrategias de inactivación disponibles para un gen codificante de proteínas pueden no ser aplicables a los loci genómicos que codifican ARNlnc. Aunque ha mejorado la anotación de genes ARNlnc (Derrien *et al.* (2012)), los límites precisos de algunos de los genes pueden permanecer ambiguos todavía, lo que puede complicar el diseño de alelos de inactivación. Una herramienta potente para aplicar a ratones nuligénicos para genes codificantes de proteínas es el reemplazo del gen diana con un indicador, tal como, por ejemplo, la secuencia codificante para β-galactosidasa o una proteína fluorescente, cuya expresión está controlada por el promotor del gen diana, informando de este modo del patrón espacial y temporal de su expresión en el ratón. Los ejemplos no limitativos de genes indicadores se proporcionan en otra parte en el presente documento.

El reemplazo de genes indicadores se ha aplicado exitosamente a ARN no codificante, tal como el locus bien estudiado Gt(ROSA)26Sor (Zambrowicz, B.P., *et al.* (1997), Proc Natl Acad Sci USA 94: 3789-3794), que codifica un ARNlnc y el gen para el gen para el ARN no codificante pequeño miR-155 (Thai, T.H., *et al.* (2007), Science 316: 604-608), pero se necesitaría desarrollar reglas para crear tales alelos para ARNlnc. A pesar de estas cualificaciones, con miles de ARNlnc identificados, ha llegado el momento de explorar la aplicación de la potencia de la tecnología de ratones nuligénicos para esta nueva clase de genes. Con este objetivo en mente, en el presente documento se describe la creación de líneas de ratones nuligénicos para veinte ARNlnc, por ejemplo, cada una portando un alelo de delección de ablación génica con, por ejemplo, un reemplazo de indicador de β-galactosidasa.

Cualquier locus de ARNlnc se puede modificar mediante los métodos y composiciones proporcionados en el presente documento. En un aspecto descrito en el presente documento, el ARNlnc es un ARN no codificante intergénico grande (ARNlinc). Los ejemplos no limitantes de genes de ARNlinc se enumeran en la Tabla 1, sin embargo, los métodos y composiciones proporcionadas en el presente documento pueden practicarse con cualquier otro ARNlnc.

La Tabla 1 enumera los 20 genes de ARNlinc en 10 cromosomas diferentes que fueron objetivo en este estudio y los 26 alelos de delección de inactivación que se crearon. Los miembros de la clase de ARN no codificante intergénico grande se eligieron para mutaciones debido a que, por definición, los genes de ARNlinc se aíslan de genes codificantes de proteínas vecinas y sus transcritos no solapan (Guttman *et al.* (2009)). Esta característica permitió el diseño de alelos de delección que tendrían menos posibilidades de interferir con la expresión de genes cercanos. Los genes de ARNlinc dirigidos se eligieron para reflejar diversos patrones de expresión (Cabili, M.N., *et al.* (2011), Genes Dev 25: 1915-1927; Khalil, A.M., *et al.* (2009), Proc Natl Acad Sci USA 106:11667-11672), con énfasis en la expresión neuronal y para su implicación potencial en el desarrollo y la regulación de la expresión génica.

Tabla 1. Alelos de delección de inactivación de ARNlinc

ARNlinc	Propiedades de delección		
	Exón de inicio de delección ¹	Tamaño (kb)	Coordenadas genómicas ²
<i>HOTAIR</i>	E1	2,3	Ch15: 102945399-102947720
	E2	0,43	Ch15: 102945399-102945826
<i>HOTTIP</i>	E1	4,8	Ch6: 52262834-52267603
	E2	2,2	Ch6: 52265374-52267603
<i>Hoxallos (HoxA11as)</i>	E3	3,5	Ch6: 52246320-52249795
	E4	3,1	Ch6: 52246643-52249795
	E5	0,70	Ch6: 52249094-52249795
<i>Pantr1 (LincRNA -Brn1-a)</i>	E1	47	Ch1: 42648175-42694815
<i>Pantr2 (LincRNA -Brn1-b)</i>	E1	6,5	Ch1: 42707143-42713698
<i>Ptgs2os2 (lincRNA-Cox2)</i>	E1	5,9	Ch1: 150159024-150164899
<i>Eldr (FabI, LincRNA-Egfr)</i>	E1	17	Ch11: 16934419-16951083
<i>Lincenc1 (LincRNA-Enc1)</i>	E2	26	Ch13: 97455710-97482249
<i>Mannr (LincRNA-Evi1)</i>	E1	32	Ch3: 29891188-29923147
<i>Fendrr (LincRNA-Foxf1)</i>	E2	19	Ch8: 121054882-121074065
<i>Halr1 (Haunt, LincRNA-HoxA1)</i>	E2	8,6	Ch6: 52106776-52115377
<i>Haglrl (Mdgf, LincRNA-HoxD3)</i>	E1	12	Ch2: 74750433-74762886
<i>Celrr (Celr, LincRNA-Insig2)</i>	E2	50	Ch1: 121087772-121137464
<i>Cmde (LincRNA-Irx5)</i>	E2	25	Ch8: 92325913-92350749

40

(continuación)

Tabla 1. Alelos de delección de inactivación de ARNlinc			
	Propiedades de delección		
<i>Kantr</i> (<i>Spasm</i> , <i>LincRNA-Jarid1c</i>)	E1	29	ChX: 152298544-152327475
<i>Pint</i> (<i>Linc-Pint</i> , <i>LincRNA-Mkln1</i>)	E2	32	Ch6: 31166026-31197846
<i>Trp53cor1</i> (<i>LincRNA-p21</i>)	E1	22	Ch17: 29057474-29079078
	E2	2,9	Ch17: 29057474-29060353
<i>Lincppara</i> (<i>LincRNA-Ppara</i>)	E1	29	Ch15: 85671665-85701064
<i>Peril</i> (<i>LincRNA-Sox2</i>)	E1	14	Ch3: 34767849-34782292
<i>Tug1</i> (<i>LincRNA-Tug1</i>)	E1	9,0	Ch11: 3639794-3648758
	E2	5,7	Ch11: 3639794-3645518
¹ A menudo los genes de ARNlinc tienen múltiples transcritos anotados. Todas las delecciones finalizan en el último exón anotado.			
² GRCm38 (GCA 000001635.3)			

La estrategia de diseño para las mutaciones de inactivación de ARNlinc estuvo guiada por dos objetivos. En primer lugar, los alelos se crearon para que indicaran con precisión la actividad de transcripción de los genes ARNlinc. Aunque hay pruebas abundantes sobre la expresión de ARNlinc específica de tejidos (Cabili *et al.* (2011)), era deseable complementar esta base de conocimiento produciendo los patrones de expresión de definición más alta ofrecida por el perfil de expresión de *lacZ*, que puede resolver la expresión de tejidos y órganos tanto espacial como temporalmente y revelar subdominios y en algunos casos, la especificidad de tipo de célula no resuelta mediante experimentos de disección de tejidos. Además, ninguno de los alelos de inactivación de ARNlinc publicados ha incorporado un indicador. Segundo, se crearon delecciones ablativas de genes que suprimieron la síntesis y la función del ARNlinc de modo que cualesquiera fenotipos asociados con las mutaciones serían informativos sobre las funciones críticas de los ARN dirigidos. Las delecciones de inactivación variaron en tamaño de aproximadamente 400 pb a 50 kb, con la mitad eliminando todos los exones anotados. Para la mayoría de los alelos restantes, la delección se inició en el segundo exón. La aplicación de los métodos de ingeniería VelociGene® (Valenzuela, D.M., *et al.* (2003a), Nat Biotechnol 21:652-659) para la construcción y uso de vectores de direccionamiento grandes basados en cromosomas artificiales bacterianos (LTVEC) fue crucial permitir la construcción de las delecciones de ablación de genes grandes, necesarias para asegurar un alelo nulo para esta nueva clase de ARN funcional grande.

Poco se sabe acerca de la relación entre estructura y función de los genes ARNlinc que pueden guiar el diseño de alelos. La experiencia con la alteración de los genes Gt(ROSA)26Sor (Zambrowicz *et al.*, (1997)) y BIC (miR-155) (Thai *et al.* (2007)), estableció que la delección e inserción después del primer exón puede producir una expresión fiable y específica de tejidos de la β -galactosidasa u otros indicadores. Esta estrategia puede, sin embargo, fallar en conseguir una mutación nula completa si el transcrito de fusión a partir del alelo modificado retiene una parte funcional del ARNlinc de la porción 5' codificada en el primer exón (Tsai, M.C., *et al.* (2010), Science 329, 689-693). Los diseños de alelos de inactivación indicados en la Tabla 1 fueron, por tanto, un compromiso entre el deseo de una mutación completamente ablativa que tuviera la probabilidad más elevada de suprimir la función de ARNlinc y el objetivo de crear un alelo que produjera un perfil de expresión de genes preciso e informativo a partir del indicador de β -galactosidasa. Por ejemplo, se hicieron dos alelos para el gen *HOTAIR*, uno que suprimió casi la secuencia codificante de ARN entera y un segundo en el que la delección se inició en el segundo exón. Ambos alelos produjeron fenotipos idénticos (descritos más adelante), pero solo el segundo funcionó como un indicador de expresión génica.

Para ARNlinc que residen muy cerca de un gen codificante de proteínas y pueden compartir un promotor divergente, el punto de inicio de delección se estableció en el segundo exón para evitar la posibilidad de alterar la transcripción del gen vecino. La Fig. 1 muestra tal ejemplo para el gen *Fendrr* (*lincRNA-Foxf1*). El diagrama muestra un ejemplo de los elementos de diseño comunes a todos los alelos: una delección dirigida de toda o la mayoría de la secuencia codificante para el ARNlinc y reemplazo con un casete que contiene una secuencia del gen *lacZ* de *E. coli* que codifica β -galactosidasa y un casete (*neo'*) que expresa neomicina fosfotransferasa para la selección de colonias de células ES resistentes a G418. Los sitios de reconocimiento de recombinasa de *LoxP* que permiten la escisión mediada por Cre antes del análisis fenotípico flanquean el casete de selección de fármacos. Como no hay marco de lectura abierto con el que fusionar la secuencia *lacZ*, cada alelo puede llevar un codón de iniciación y una secuencia consenso Kozak (Kozak, M. (1987), Nucleic Acids Res 15, 8125-8148) para la traducción eficaz del indicador de β -galactosidasa. Son ejemplos no limitativos de secuencias consenso Kozak A/GCCRCCATGG (SEQ ID NO: 1) y GCCGCCRCCATGG (SEQ ID NO: 2), en donde R es A o G.

Los vectores de direccionamiento LTVEC se introdujeron en células ES y se cribaron para los clones dirigidos correctamente mediante el método de pérdida de alelo (Frendewey, D., *et al.* (2010), Methods Enzymol 476, 295-307).

El método VelociMouse® (Poueymirou, W.T., *et al.* (2007), Nat Biotechnol 25, 91-99) se aplicó a la inyección de la etapa embrionaria de 8 células para convertir las células ES dirigidas en ratones heterocigóticos de generación F0

derivados de células ES listos para el perfil de expresión de *lacZ* o la crianza hacia la homocigosis. Se proporcionan detalles adicionales de los métodos para generar animales nuligénicos de ARNlnc en los Ejemplos 1-13 en otra parte en el presente documento.

5 ii. Perfil de expresión del indicador

10 Como se describe en otra parte en el presente documento, la modificación genética del locus de ARNlnc puede comprender un reemplazo de o una inserción/adición en el locus de ARNlnc o una porción del mismo con un inserto de ácido nucleico. En algunos casos, el inserto de ácido nucleico comprende un gen indicador. En un aspecto descrito en el presente documento, el gen indicador se posiciona en el locus de ARNlnc en unión operativa con el promotor de ARNlnc endógeno. Tal modificación permite la expresión del gen indicador impulsada por el promotor de ARNlnc endógeno. Como alternativa, el gen indicador no se sitúa en unión operativa con el promotor de ARNlnc endógeno.

15 Se puede utilizar cualquier indicador (o fracción detectable) en los métodos y composiciones proporcionados en el presente documento. Los ejemplos no limitantes de indicadores incluyen, por ejemplo, β -galactosidasa (codificada por el gen *lacZ*), proteína fluorescente verde (GFP; del inglés, Green Fluorescent Protein), proteína fluorescente verde potenciada (eGFP; del inglés, enhanced GFP), mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrina, Venus, YPet, proteína fluorescente amarilla potenciada (EYFP), Esmeralda, CyPet, proteína fluorescente cian (CFP), Cerulean, T-Zafiro, luciferasa, fosfatasa alcalina o una combinación de los mismos.

20 La siguiente descripción es un ejemplo no limitativo que utiliza un gen indicador *lacZ* que codifica una β -galactosidasa. Los métodos y composiciones descritos en el presente documento pueden realizarse con cualquier otro gen indicador.

25 Para investigar los patrones de expresión de los 20 genes de ARNlnc dirigidos se aplicó una tinción de X-gal para actividad de β -galactosidasa sobre embriones completos o tejidos y órganos montados completos de ratones adultos. Los genes de ARNlnc dirigidos mostraron una diversidad de patrones de expresión de un único gen indicador (Tabla 2 en la Fig. 11), que representa la mayoría de los sistemas de órganos principales y tipos de tejido. Los patrones de expresión del indicador indican que la mayoría de los ARNlnc se transcriben en múltiples tejidos de adultos, con un gen, *Pint*, que muestra una expresión ubicua en todos los tejidos examinados. Para aproximadamente un tercio de los genes ARNlnc, la expresión se restringió a un solo órgano, por ejemplo, cerebro para *Pantr2*, *Kantr* y *Hagl*, pulmones para *Mannr* y *Fendrr*, el sistema urogenital para *Eldr* y la caja torácica para *Halr1*. Tres de los genes ARNlnc que incluyen *HOTAIR*, *Ptgs2os2*, y *Hagl* no mostraron expresión en ningún tejido de adulto.

35 La expresión embrionaria parece ser una característica común de los ARNlnc. El examen de la expresión del indicador de β -galactosidasa en embriones heterocigóticos el día o alrededor del día embrionario 12,5 (E12,5) reveló diversos patrones específicos para todos los 20 genes de ARNlnc dirigidos (Tabla 2 en la Fig. 11, Fig. 2A). Los perfiles de expresión variaron desde ubicuos (*Tug1*) a altamente restringidos, tales como epidérmico para *Eldr*, placoda de bigotes para *Trp53cor1* (Fig. 9) o los brotes mamarios para *Lincenc1* (Fig. 9). Los patrones espaciotemporales observados en los diferentes alcances de la expresión de brote de extremidades y cola para *HOTTIP* y *Hoxallos* son muy similares a aquellos indicados para los genes codificantes de proteínas adyacentes en el grupo HoxA (Hostikka, S.L. y Capecchi, M.R. (1998), *Mech Dev* 70:133-145; Lu, P., *et al.* (2008), *Development* 135:1395-1405). La expresión de *HOTAIR* en el brote posterior de cola y tubérculo genital que se observó para el indicador de β -galactosidasa fue idéntica a la determinada mediante hibridación *in situ* (Schorderet, P. y Duboule, D. (2011), *PLoS Genet* 7:e1002071). El análisis de tinción de β -galactosidasa en diferentes momentos del desarrollo embrionario mostró que para algunos de los ARNlnc la expresión comenzó de forma temprana en un sitio restringido y a continuación se extendió más allá este locus inicial en etapas posteriores (Fig. 2B), de nuevo, recuerda a la expresión de proteínas de Hox (Nagy, A. (2003) "Manipulating the mouse embryo: a laboratory manual", Cold Spring Harbor, Nueva York, Cold Spring Harbor Laboratory Press). Por ejemplo, la expresión de los genes *HOTTIP* y *Hoxallos* comenzó en el extremo posterior del embrión de E9,5 y a continuación, se extendió en los brotes de extremidades en momentos posteriores. De forma análoga, la expresión inicial de *Celrr* en un sitio cerca del extremo anterior de los embriones en E9,5 se mantuvo y se expandió a toda la longitud de la médula espinal en los siguientes dos días.

50 En coherencia con la expresión cerebral frecuente observada entre los ARNlnc específicos de tejidos humanos (Derrien *et al.* (2012)), la mitad de los veinte genes de ARNlnc de ratón dirigidos son activos transcripcionalmente en el cerebro. Como con la expresión de ARNlnc embrionaria, los patrones cerebrales (Fig. 3) resultaron únicos y variaron desde ubicuos (*lincppara* y *Pint*) a altamente restringido (*Peril*, *Crnde* y *Kantr*). Entre los 20 genes de ARNlnc que se dirigieron, solo *Pint* mostró un patrón de expresión corporal completo, que estuvo principalmente restringido a la vida posnatal. Exclusivo de *Pint*, se observó un incremento de su expresión con la edad (Fig. 4). En neonatos de 3 días, la actividad transcripcional de *Pint* es baja (cerebro) o indetectable (músculo de la caja torácica), pero a continuación aparece gradualmente en ratones de 3 semanas de edad y se vuelve fuerte y ubicua a las 8 semanas de edad. Aunque la fuerza y el momento de expresión de *Pint* varía entre los distintos órganos y tejidos, la tendencia general es un incremento constante en la expresión después del nacimiento a una meseta en la edad adulta. Este patrón de expresión dinámico relacionado con la edad es nuevo; los inventores no han observado un perfil similar en experimentos de perfilado de *lacZ* para cientos de genes de inactivación codificantes de proteínas.

65 iii. Fenotipos

La modificación genética de loci de ARNlnc puede dar como resultado diversos fenotipos en los animales no humanos proporcionados en el presente documento. Tales fenotipos pueden incluir, por ejemplo, un fenotipo asociado a envejecimiento prematuro, defectos en el desarrollo de diversos órganos incluyendo cerebro, esqueleto, músculo o pulmón, defectos en el desarrollo embrionario, letalidad perinatal o embrionaria, pérdida de pelo, detención del crecimiento prematura, ciflordosis o postura anormal.

En un aspecto descrito en el presente documento, un animal no humano que comprende al menos un ARN no codificante largo (ARNlnc) modificado como se describe en el presente documento se caracteriza por tener uno o más de los siguientes fenotipos: (a) un fenotipo asociado a envejecimiento prematuro; (b) letalidad perinatal; (c) un defecto en el desarrollo pulmonar; (d) una malformación morfológica en la cola y en las extremidades traseras; (e) una pérdida de masa muscular en uno o más tejidos; (f) un defecto en el desarrollo cerebral; o (j) una combinación de los mismos de cualquiera de (a)-(f).

En un aspecto descrito en el presente documento, la modificación genética del locus de ARNlnc da como resultado la letalidad. En algunos casos la modificación del locus de ARNlnc es letal embrionario. En un aspecto descrito en el presente documento, la modificación del locus de ARNlnc es letal perinatal. En un aspecto descrito en el presente documento, una alteración o inactivación del ARNlnc *Fendrr* o ARNlnc *Aperil* da como resultado la letalidad perinatal. En otro aspecto descrito en el presente documento, una alteración o inactivación de *Haglrl* da como resultado la letalidad perinatal.

En un aspecto descrito en el presente documento, la modificación genética del locus de ARNlnc da como resultado un fenotipo asociado a envejecimiento prematuro. En tales animales, los signos de envejecimiento prematuro pueden incluir, por ejemplo, ganancia de peso lenta, peso corporal anterior estancado, detención del crecimiento prematura, ciflordosis a las 12 semanas de edad aproximadamente, ciflordosis grave a las 26 semanas de edad aproximadamente, pérdida de pelaje a los seis meses aproximadamente, pérdida de fuerza muscular de las patas traseras a los 6 meses aproximadamente o una combinación de los mismos. En un aspecto descrito en el presente documento, la modificación genética que da como resultado un fenotipo asociado a envejecimiento prematuro es una alteración o inactivación de *Pint*. En un aspecto descrito en el presente documento, el ARNlnc es *Pint* y el animal no humano se caracteriza por un fenotipo asociado a envejecimiento prematuro que comprende: (a) una tasa de crecimiento más lenta que la de un control de tipo silvestre; (b) un declive en la fuerza muscular; (c) fibrosis; (d) un contenido de grasa corporal más bajo que el del control de tipo silvestre; (e) una densidad mineral ósea del fémur y una masa ósea más baja que la del control de tipo silvestre; (f) una masa muscular disminuida en comparación con la del control de tipo silvestre; (g) una disminución en la longevidad media; (h) ciflordosis; (i) atrofia orgánica; o (j) una combinación de los mismos de cualquiera de (a)-(i).

La modificación genética de pérdida de función del locus de ARNlnc puede también dar como resultado un defecto en el desarrollo cerebral. En un aspecto descrito en el presente documento, la modificación genética que da como resultado un fenotipo asociado a envejecimiento prematuro es una alteración o inactivación de *Pantr2*, *Kantr*, *Peril*, *Celrr*, *Pantr1*, *Crnde*, *lincnc1*, *Pint*, *lincppara* o *Tug1*. En un aspecto específico descrito en el presente documento, el ARNlnc es *Pantr2*. En otro aspecto específico descrito en el presente documento, el ARNlnc es *Pint*.

Se realizó un análisis fenotípico de los diversos ejemplos de ratones nuligénicos para ARNlnc proporcionado en el presente documento y se describe en el presente documento más adelante.

El sorprendente incremento relacionado con la edad en la expresión de *Pint* corporal completa revelado por el perfilado de *lacZ* (Fig. 4) sugirió que *Pint* puede tener un papel homeostático global en el mantenimiento de la salud normal según envejecen los ratones. Para analizar esta hipótesis el alelo de inactivación se crió para homocigosis y se siguieron los ratones de tipo silvestre (WT), heterocigóticos (Het) y nuligénicos (KO) desde el nacimiento hasta las 26 semanas de edad y se examinaron la tasa de crecimiento y signos evidentes cualesquiera de mala salud o defecto. Los ratones KO para *Pint* ganaron peso con la edad en una tasa más lenta y alcanzaron la meseta de peso corporal más pronto y a un peso corporal significativamente más bajo que los ratones WT (Fig. 5A y B), implicando una detención del crecimiento prematura. Los ratones KO y Het tanto machos como hembras mostraron el fenotipo de crecimiento lento, pero fue más pronunciado en los machos. Las imágenes esqueléticas mediante análisis de microCT (microtomografía computerizada) de ratones individuales según envejecían revelaron la aparición de ciflordosis en aproximadamente un 70 % de los ratones KO para *Pint* a las 12 semanas de edad, con cerca de un 90 % de los ratones KO de 26 semanas de edad mostrando ciflordosis grave (Fig. 5C y D). En cambio, solo un 10 a un 20 % de los ratones WT de 26 semanas de edad mostraron una ligera ciflordosis relacionada con la edad. No apareció una ciflordosis significativa en los ratones Het *Pint* hasta las 26 semanas de edad, indicando una haploinsuficiencia dependiente de la edad inusual para *Pint*. Una pérdida de pelaje relacionada con la edad en ratones KO de 6 meses de edad se observó también que fue más grave en hembras (5 de 10 en KO) que en machos (2 de 9) y se observó solo en un Het y en ninguno de los ratones WT de la misma edad. Un fenotipo menos grave, comportamiento de juntado de las patas traseras cuando se suspendieron de la cola, se percibió en dos tercios aproximadamente de los ratones KO para *Pint* de 6 meses de edad (60 % de hembras, 67 % de machos) en comparación con aproximadamente el 20 % de ratones WT de la misma edad. Este fenotipo pudo indicar una pérdida relacionada con la edad de fuerza muscular de patas traseras (véase la Fig. 8 para otro ejemplo de este fenotipo en la línea nuligénica para *HOTTIP*). El

espectro de defectos asociados a mutaciones en los ratones nuligénicos para *Pint* sugiere un fenotipo asociado a envejecimiento prematuro.

De las 19 líneas de ratones nuligénicos para ARNlinc que se criaron para homocigosis, dos (11 %), *Peril* y *Fendrr*, mostraron una letalidad perinatal (L.A. Goff *et al.*, no publicado). Recientemente, se ha informado de una inactivación del gen *Fendrr* (Grote *et al.* (2013)). El alelo consistió en la inserción de un elemento de parada transcripcional en el primer exón. En embriones que llevan una mutación homocigótica de *Fendrr*, Grote *et al.* (2013) observaron letalidad alrededor de E13,75 que estaba asociada con una hernia umbilical prominente, una reducción en el grosor de la pared corporal ventral y un defecto cardíaco que provocó acumulación de sangre en el atrio derecho. No se observó ninguno de estos fenotipos en la línea de inactivación de *Fendrr* descrita en el presente documento, que tiene una delección de 26 kb desde el exón 2 al último exón anotado (Fig. 1). La tinción de X-gal de embriones de E12,5 mostró una expresión de *lacZ* en el proceso frontonasal, tracto respiratorio superior, pulmones y en la región AGM (aorta gónada mesonefros) posterior (Fig. 6A) que fue idéntica en embriones tanto heterocigóticos (no mostrado) como homocigóticos. Una mirada aislada a los pulmones en desarrollo en la E13,5 reveló defectos en los embriones nuligénicos: los lóbulos pulmonares parecieron colapsados, globulares y desorganizados (Fig. 6B). Los ratones homocigóticos para el alelo de delección inactivación del gen *Fendrr* descrito en el presente documento sobrevivieron al nacimiento pero sucumbieron poco después por problemas respiratorios evidentes. El fenotipo letal perinatal mutante *Fendrr* fue idéntico en ratones sobre 2 antecedentes genéticos diferentes: el antecedente híbrido C57B16NTac/129S6SvEvTac informado aquí y en ratones adicionalmente retrocruzados en un antecedente de C57BL/6 en un programa de cría separado.

Se observaron también varios fenotipos moderados entre los ratones nuligénicos para ARNlinc que sobrevivieron a la edad adulta y mostraron una segregación mendeliana normal de los alelos mutantes (L.A. Goff *et al.*, no publicado). Entre estos, hay una fuerte correlación entre las expresiones de *lacZ* y los fenotipos detectables en varias inactivaciones de genes de ARNlinc tales como *Pantr2*, *HOTAIR* y *HOTTIP*. Los ratones homocigóticos para la ablación del gen *Pantr2* (una delección de secuencia codificante completa de 6,5 kb, Tabla 1) muestran un desarrollo inadecuado del córtex cerebral (fuerte tinción en el SNC), con cifras reducidas de células progenitoras intermedias y defectos en la migración neuronal que afecta al desarrollo de neuronas de proyección de capas superiores (L.A. Goff *et al.*, no publicado). Las delecciones de los genes *HOTAIR* y *HOTTIP* (ablaciones génicas completas o parciales, Tabla 1) causaron malformaciones morfológicas totalmente penetrantes. Se observó en ratones KO para *HOTAIR* una transformación homeótica aparente de la 4ª vértebra caudal, que parece ser anatómicamente similar a la 3ª vértebra caudal (Fig. 7) (tinción de *lacZ* transitoria específica en el brote de la cola). Los ratones KO para *HOTTIP* (que tuvieron una tinción positiva en los brotes de extremidades embrionarias) mostraron una postura anormal en comparación con los compañeros de camada de tipo silvestre cuando se suspendieron de la cola (Fig. 8A). Esta anomalía de comportamiento se acompañó de una pérdida en la resistencia de prensión como se midió mediante una prueba en la que se desafiaba a los ratones a permanecer suspendidos en una jaula de alambre invertida. El tipo silvestre y los mutantes Het *HOTTIP* se mantuvieron durante aproximadamente un minuto, mientras que sus compañeros de camada KO liberaron su prensión en 10-20 segundos (Fig. 8B). Esta reducción evidente de la fuerza de prensión se asocia con una pérdida de masa muscular del gastrocnemio, pero no del tibial anterior o del cuádriceps (Fig. 8C). Se observó un 40 % aproximadamente de reducción en las cifras de fibras en el gastrocnemio, pero no una reducción en el tamaño promedio (Fig. 8D y E). Además de los defectos musculares en los ratones nuligénicos para *HOTTIP*, se encontró también una malformación esquelética: un acortamiento en la longitud del hueso calcáneo de la pata trasera (Fig. 10).

En los últimos años se ha visto una explosión en la comprensión del componente no codificante de proteínas del genoma, especialmente en mamíferos. Además de las clases de ARN funcionales no codificantes conocidos durante décadas, ARN ribosomales, de transferencia, nucleares pequeños, nucleolares pequeños, citoplasmáticos pequeños y los componentes de ARN de la ARNasa P, ARNasa MRP y enzimas telomerasas y los descubiertos más recientemente microARN y los ARNpi asociados a PIWI, se pueden incluir ahora al menos 10.000 miembros de la clase ARN no codificante largo (Carninci *et al.* (2005); Derrien *et al.* (2012); Djebali *et al.* (2012); Guttman *et al.* (2009); Kapranov *et al.* (2007)). Según vamos entendiendo la presencia genómica y la expresión de genes ARNlinc, el siguiente objetivo es descubrir sus funciones biológicas. Como una primera etapa para empezar a abordar este reto, se ha aplicado la tecnología de ratones nuligénicos, la herramienta más potente para la determinación de la función génica en mamíferos, para crear un recurso de líneas de ratones nuligénicos para 20 genes de ARNlinc seleccionados por su expresión predominantemente de linaje neuronal y su función esperada en el desarrollo.

Dadas las relaciones estructura-función desconocidas para los ARNlinc, fue crucial en este estudio inicial crear alelos de inactivación con delecciones que suprimieron la mayoría si no todo el potencial codificante de ARNlinc, para asegurar que cualesquiera fenotipos que se observaron fueron el resultado de alelos completamente nulos. La anotación ambigua y complicada de muchos loci de ARNlinc, con múltiples transcritos informados, tal vez generados mediante corte y empalme alternativo o sitios de inicio de la transcripción, se añade a la dificultad del diseño de alelos de inactivación y haría difícil la construcción de alelos condicionales que eviten el riesgo de efectos hipomórficos. La nueva comprensión de las características moleculares importantes para la función de ARNlinc debe informar del diseño de la siguiente generación de alelos de ARNlinc con modificaciones dirigidas de forma más precisa de secuencias críticas para funcionar y permitir también estrategias condicionadas avanzadas y flexibles.

Un objetivo de la investigación de inactivación de ARNlinc descrito en el presente documento fue crear alelos que además de abolir una función también reportó el patrón espacio temporal de expresión del gen. A pesar de no tener

un marco de lectura abierto como una guía, se diseñaron alelos satisfactoriamente que informaron de expresión génica de todos los 20 genes dirigidos. Uno de los alelos que no produjo expresión de *lacZ* en la etapa adulta fue *Ptgs2os2* (véanse la Fig. 2A y la Fig. 9 para la expresión embrionaria), que se conoce por ser uno de los ARNlinc más fuertemente inducidos por señales inflamatorias (Carpenter, S., *et al.* (2013), *Scienceexpress* 01 de agosto de 2013; Guttman *et al.* (2009)). En esta investigación no se realizaron experimentos de desafío, pero la línea de inactivación *Ptgs2os2* debe demostrar ser un recurso valioso para estudios de cómo una expresión de ARNlinc responde a una infección u otros ataques inflamatorios y qué papel biológico juega en el proceso.

La variedad y especificidad de los patrones de expresión génica que se observaron recordaron a aquellos vistos con alelos indicadores para genes codificantes de proteínas. La expresión embrionaria fue una característica compartida por casi todos los genes ARNlinc examinados. Esto podría señalar un papel común para los ARNlinc en la regulación de eventos clave en el desarrollo. Se observaron patrones de cambio espaciotemporales durante el desarrollo embrionario, como las proteínas HOX (Fig. 2B), expresión exquisitamente específica, tal como la tinción de placoda de bigotes para *Trp53cor1* y la expresión de brotes mamarios para *Lincenc1* (Fig. 2A y Fig. 9), expresión ubicua en tejidos adultos, como para *Pint* (Fig. 4) y cambios temporales en los patrones de expresión, tales como cambios cualitativos observados para *Celrr* (Fig. 2B) en el desarrollo embrionario o el incremento cuantitativo nuevo en la expresión global con la edad para *Pint* (Fig. 4). Como muchos de los genes ARNlinc elegidos para esta investigación eran conocidos por estar expresados en linajes celulares neuronales, se observó la expresión de indicadores específicos de cerebro (Fig. 3), pero el perfilado de *lacZ* proporcionó una mayor resolución con información biológica más rica que los ensayos basados en células o tejidos diseccionados.

Entre los fenotipos observados entre las 19 líneas de inactivación de ARNlinc criados para homocigosis, se observó letalidad dos veces (11 %), una frecuencia que probablemente es más baja de lo que sería de esperar para inactivaciones de 20 genes codificantes de proteínas elegidos aleatoriamente. La tasa algo baja de letalidad junto con la sutileza de los fenotipos no letales observados para *HOTAIR* (Fig. 6), *HOTTIP* (Fig. 7 y 8), *Pantr2* (L.A. Goff *et al.*, no publicado) y otros (M. Sauvageau *et al.*, no publicado) y la expresión embrionaria frecuente sugiere que los ARNlinc podrían amortiguar o modular la expresión génica u otros procesos más que cumplir funciones individuales, funciones críticas. En este sentido los ARNlinc podrían ser similares a sus primos no codificantes más pequeños, los miARN, en que podrían compartir objetivos redundantes y solapantes y funciones con otros ARNlinc funcionales.

Un objetivo de este trabajo fue generar un recurso de líneas de ratones nuligénicos para ARNlinc con una estrategia alélica común y una capacidad indicadora funcional que pudiera servir como los sujetos de estudios de expresión y fenotípicos más en profundidad. La adición del casete *LacZ* en cada caso permitió una alteración simultánea de la función génica y estudiar la regulación de los patrones de expresión de ARNlinc mediante tinción X-Gal. Estos estudios revelan patrones dinámicos espaciales y temporales de expresión de ARNlinc durante la embriogénesis de ratones y a lo largo de la vida adulta, dan una idea significativa de las propiedades de regulación/función de esta nueva clase de moléculas *in vivo* y localizan regiones donde la función de estos genes se puede observar. Esta investigación podría servir como modelo para un proyecto a gran escala para mutar todos los miembros de la clase de ARNlinc, similar a lo que se ha conseguido mediante el consorcio para ratones nuligénicos internacional (en inglés, International Knockout Mouse Consortium) para genes codificantes de proteínas (Bradley *et al.* (2012)).

II. Métodos para modificar un locus de ARNlinc en animales no humanos

Los métodos para modificar genéticamente un locus de ARNlinc en animales no humanos, células, tejidos o embriones, se proporcionados en el presente documento.

Cualquier locus de ARNlinc se puede modificar mediante los métodos proporcionados en el presente documento. Los ejemplos no limitantes de genes ARNlinc comprenden *HOTAIR*, *HOTTIP*, *Hoxallos* (denominado anteriormente *HoxA11as*), *Pantr1* (denominado anteriormente *lincRNA-Brn1-a*), *Pantr2* (denominado anteriormente *lincRNA-Brn1-b*), *Ptgs2os2* (denominado anteriormente *lincRNA-Cox2*), *Eldr* (denominado anteriormente *Fabl* y *lincRNA-Egfr*), *lincenc1* (denominado anteriormente *lincRNA-Enc1*), *Mannr* (denominado anteriormente *lincRNA-Evil*), *Fendrr* (denominado anteriormente *lincRNA-Foxf1*), *Halr1* (denominado anteriormente *Haunt* y *lincRNA-HoxA1*), *Haglr* (denominado anteriormente *Mdgt* y *lincRNA-HoxD3*), *Celrr* (denominado anteriormente *Celr* y *lincRNA-Insig2*), *Crnde* (denominado anteriormente *lincRNA-Irx5*), *Kantr* (denominado anteriormente *Spasm* y *lincRNA-Jarid1c*), *Pint* (denominado anteriormente *line-Pint* y *lincRNA-Mklnl*), *Trp53cor1* (denominado anteriormente *lincRNA-p21*), *lincppara* (denominado anteriormente *lincRNA-Ppara*), *Peril* (denominado anteriormente *lincRNA-Sox2*), *Tug1* (denominado anteriormente *lincRNA-Tug1*) o una combinación de los mismos.

En un aspecto descrito en el presente documento, se proporciona un método para modificar un locus de interés de ARNlinc en una célula pluripotente. Tal método comprende (a) introducir en la célula pluripotente un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado con brazos de homología 5' y 3' que pueden sufrir recombinación homóloga con el locus de ARNlinc; y (b) identificar una célula pluripotente modificada que comprende una modificación genética dirigida al locus de ARNlinc. En tales métodos, la modificación genética da como resultado la pérdida de función del ARNlinc. En un aspecto descrito en el presente documento, la célula pluripotente es una célula embrionaria de ratón o de rata. En otro aspecto descrito en el presente documento, la célula pluripotente es una célula humana iPS.

A. Vectores de direccionamiento e insertos de ácidos nucleicos

5 Se proporcionan adicionalmente vectores de direccionamiento o constructos de direccionamiento para emplearse en los métodos para hacer los animales no humanos, células, tejidos o embriones genéticamente modificados proporcionados en el presente documento.

10 En un aspecto descrito en el presente documento, se proporciona un vector de direccionamiento que comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3' que pueden someterse a recombinación homóloga con un locus de ARNInc de interés.

15 Los vectores de direccionamiento y los ejemplos de componentes de los vectores de direccionamiento (es decir, insertos de ácidos nucleicos, polinucleótidos de interés, casetes de expresión, etc.) se describen en detalle en el presente documento más adelante.

i. Inserto de ácido nucleico

20 El "inserto de ácido nucleico" o "inserto de polinucleótido" comprende un segmento de ADN que se desea integrar en el locus diana. En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende uno o más polinucleótidos de interés. En otros aspectos descritos en el presente documento, el inserto de ácido nucleico puede comprender uno o más casetes de expresión. Un casete de expresión dado puede comprender un polinucleótido de interés, un polinucleótido que codifica un marcador de selección y/o un gen indicador junto con los diversos componentes reguladores que influyen en la expresión.

25 Cualquier polinucleótido de interés puede estar contenido en los diversos insertos de polinucleótidos y por tanto integrado en el locus genómico diana. Los métodos desvelados en el presente documento, proporcionan al menos 1, 2, 3, 4, 5, 6 o más polinucleótidos de interés para que se integren en el locus genómico de ARNInc dirigido de interés.

30 En un aspecto descrito en el presente documento, el polinucleótido de interés contenido en el inserto de ácido nucleico codifica un indicador. En otro aspecto descrito en el presente documento, el polinucleótido de interés codifica un marcador seleccionable.

35 En un aspecto descrito en el presente documento, el polinucleótido de interés puede estar flanqueado por sitios de recombinación específica de sitio. En un aspecto específico descrito en el presente documento, los sitios de recombinación específica de sitio flanquean un segmento que codifica un indicador y/o un segmento que codifica un marcador seleccionable.

40 Los ejemplos no limitativos de polinucleótidos de interés, incluyendo marcadores de selección y genes indicadores que se pueden incluir dentro del inserto de ácido nucleico se analiza en detalle en otra parte en el presente documento.

45 El polinucleótido de interés dentro del inserto de polinucleótido cuando se integra en el locus de ARNInc diana puede introducir una o más modificaciones genéticas en la célula. La modificación genética puede comprender una delección de una secuencia de ácido nucleico y/o la adición de un polinucleótido exógeno o heterólogo u ortólogo en el locus genómico diana. En un aspecto descrito en el presente documento, la modificación genética comprende un reemplazo de una secuencia de ácido nucleico endógena con un polinucleótido exógeno de interés en el locus genómico diana. Por tanto, los métodos proporcionados en el presente documento para la generación de una modificación genética comprenden una inactivación, una delección, una inserción, un reemplazo ("activación"; del inglés, "knock-in"), una mutación puntual, un intercambio de dominios, un intercambio de exones, un intercambio de intrones, un intercambio de secuencia reguladora, un intercambio génico o una combinación de los mismos en un locus de ARNInc diana. Tales modificaciones pueden ocurrir tras la integración del primer, segundo, tercero, cuarto, quinto, sexto, séptimo o cualesquier insertos de polinucleótidos subsiguientes en el locus genómico diana.

55 El polinucleótido de interés dentro del inserto de polinucleótido y/o integrado en el locus genómico diana puede comprender una secuencia que es nativa y homóloga para la célula en la que se introduce; el polinucleótido de interés puede ser heterólogo para la célula en la que se introduce; el polinucleótido de interés puede ser exógeno para la célula en la que se introduce; el polinucleótido de interés puede ser ortólogo para la célula en la que se introduce; o el polinucleótido de interés puede ser de una especie diferente a la célula en la que se introduce. El término "homólogo" en referencia a una secuencia es una secuencia que es nativa para la célula. El término "heterólogo" en referencia a una secuencia es una secuencia que se origina de una especie ajena o si es de la misma especie, está sustancialmente modificada con respecto a su forma nativa en composición y/o locus genómico por intervención humana deliberada. El término "exógeno" en referencia a una secuencia es una secuencia que se origina de una especie ajena. El término "ortólogo" es un polinucleótido de una especie que es funcionalmente equivalente a una secuencia de referencia conocida en otra especie (es decir, una variante de especie). El polinucleótido de interés puede ser de cualquier organismo de interés, incluyendo, pero sin limitarse a, un procarionta, un eucariota, un no humano, un roedor, un hámster, un ratón, una rata, un ser humano, un mono, un ave, un mamífero agrícola o un mamífero no agrícola. El polinucleótido de interés puede adicionalmente comprender una región codificante, una región no codificante, una

región reguladora o un ADN genómico. Por tanto, el 1^{er}, 2^o, 3^{er}, 4^o, 5^o, 6^o, 7^o y/o cualesquier insertos de polinucleótidos subsiguientes pueden comprender tales secuencias.

5 En un aspecto descrito en el presente documento, el polinucleótido de interés puede variar de aproximadamente 500 nucleótidos a aproximadamente 200kb como se ha descrito anteriormente. El polinucleótido de interés puede ser de aproximadamente 500 nucleótidos a aproximadamente 5kb, de aproximadamente 5kb a aproximadamente 200kb, de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 60kb a aproximadamente 70kb, de
10 aproximadamente 80kb a aproximadamente 90kb, de aproximadamente 90kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 110kb, de aproximadamente 120kb a aproximadamente 130kb, de aproximadamente 130kb a aproximadamente 140kb, de aproximadamente 140kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 160kb, de aproximadamente 160kb a aproximadamente 170kb, de aproximadamente 170kb a aproximadamente 180kb, de aproximadamente 180kb a aproximadamente 190kb o de
15 aproximadamente 190kb a aproximadamente 200kb.

El polinucleótido de interés dentro del inserto de polinucleótido y/o insertado en el locus genómico diana puede codificar un polipéptido, puede codificar un ARN, puede codificar un miARN o puede comprender cualesquiera regiones reguladoras o regiones no codificantes de interés, incluyendo, por ejemplo, una secuencia reguladora, una
20 secuencia promotora, una secuencia potenciadora, una secuencia de unión al represor transcripcional, un segmento consenso Kozak, un codón de inicio o una delección de una secuencia no codificante de proteínas, pero no comprende una delección de una secuencia codificante de proteínas. Además, el polinucleótido de interés dentro del inserto de polinucleótido y/o insertado en el locus genómico diana puede codificar una proteína expresada en el sistema nervioso, el sistema esquelético, el sistema digestivo, el sistema circulatorio, el sistema muscular, el sistema respiratorio, el
25 sistema cardiovascular, el sistema linfático, el sistema endocrino, el sistema urinario, el sistema reproductor o una combinación de los mismos.

En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende un alelo activado de al menos un exón de un gen endógeno. En un aspecto descrito en el presente documento, el inserto de ácido nucleico
30 comprende un alelo activado del gen endógeno completo (es decir, "activación de gen intercambiado").

En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende un elemento regulador, que incluye por ejemplo, un promotor, un potenciador o un elemento de unión al represor transcripcional.

35 En aspectos adicionales descritos en el presente documento, el inserto de ácido nucleico comprende un alelo condicional. En un aspecto descrito en el presente documento, el alelo condicional es un alelo multifuncional, como se describe en el documento US 2011/0104799. En aspectos específicos descritos en el presente documento, el alelo condicional comprende: (a) una secuencia de actuación en orientación sentido respecto a la transcripción de un gen diana y un casete de selección de fármacos en orientación sentido o antisentido; (b) una secuencia nucleotídica de interés (NSI; del inglés, nucleotide sequence of interest) en orientación antisentido y un condicional por módulo de
40 inversión (COIN, que utiliza un intrón que divide exones y un módulo invertible similar a una trampa genética; véase, por ejemplo, el documento US 2011/0104799); y (c) unidades recombinables que se recombinan tras la exposición a una primera recombinasa para formar un alelo condicional que (i) carece de la secuencia de actuación y del DSC y (ii) contiene el NSI en orientación sentido y el COIN en orientación antisentido.

45 En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende una modificación genética en una secuencia codificante. En un aspecto descrito en el presente documento, la modificación genética comprende una mutación de delección de una secuencia codificante. En un aspecto descrito en el presente documento, la modificación genética comprende una fusión de dos secuencias codificantes endógenas.
50

En un aspecto descrito en el presente documento, la modificación genética comprende una delección de una secuencia no codificante de proteínas, pero no comprende una delección de una secuencia codificante de proteínas. En un aspecto descrito en el presente documento, la delección de la secuencia no codificante de proteínas comprende una delección de un locus de ARNlnc o una porción del mismo. En un aspecto descrito en el presente documento, la delección de la
55 secuencia no codificante de proteínas comprende una delección de un elemento regulador. En un aspecto descrito en el presente documento, la modificación genética comprende una delección de un promotor. En un aspecto descrito en el presente documento, la modificación genética comprende una adición de un promotor o un elemento regulador. En un aspecto descrito en el presente documento, la modificación genética comprende un reemplazo de un promotor o un elemento regulador.
60

65 En un aspecto descrito en el presente documento, la secuencia de ácido nucleico del vector de direccionamiento puede comprender un polinucleótido que cuando se integra en el genoma producirá una modificación genética de una región del mamífero, animal no humano o un locus de ARNlnc de mamífero no humano, en donde la modificación genética en el locus de ARNlnc da como resultado una pérdida de función del ARNlnc. En un aspecto descrito en el presente documento, se genera una inactivación de ARNlnc ("alelo nulo"). En otro aspecto descrito en el presente documento, se genera una alteración en el locus de ARNlnc.

En aspectos adicionales descritos en el presente documento, el inserto de ácido nucleico da como resultado el reemplazo de una porción del locus de ARNlnc de mamífero, de animal no humano o de mamífero no humano, con una secuencia de inserto de ácido nucleico. En un aspecto descrito en el presente documento, la secuencia del inserto de ácido nucleico es una secuencia de ácido nucleico indicadora.

El inserto de polinucleótido dado y la región correspondiente del locus de mamífero, de no humano o de mamífero no humano a reemplazar puede ser una región no codificante, una región codificante, un intrón, un exón, una región no traducida, una región reguladora, un promotor o un potenciador o cualquier combinación de los mismos. Por otra parte, el inserto de polinucleótido dado y/o la región del locus de mamífero, de no humano o de mamífero no humano a ser eliminada puede ser de cualquier longitud deseada, incluyendo por ejemplo, entre 10-100 nucleótidos de longitud, 100-500 nucleótidos de longitud, 500-1kb nucleótidos de longitud, 1kb a 1,5kb nucleótidos de longitud, 1,5kb a 2kb nucleótidos de longitud, 2kb a 2,5kb nucleótidos de longitud, 2,5kb a 3kb nucleótidos de longitud, 3kb a 5kb nucleótidos de longitud, 5kb a 8kb nucleótidos de longitud, 8kb a 10kb nucleótidos de longitud o más. En otros casos, el tamaño de la inserción o reemplazo es de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb. En otros aspectos descritos en el presente documento, el inserto de polinucleótido dado y/o la región del locus de mamífero, de célula humana o de mamífero no humano a ser eliminada es al menos de 100, 200, 300, 400, 500, 600, 700, 800, o 900 nucleótidos o al menos de 1 kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, 9kb, 10kb, 11kb, 12kb, 13kb, 14kb, 15kb, 16kb, 17kb, 18kb, 19kb, 20kb, 25kb, 30kb, 35kb, 40kb, 45kb, 50kb o más.

En un aspecto descrito en el presente documento, el inserto de ácido nucleico se inserta en el locus de ARNlnc de interés de manera que está unido operativamente al promotor de ARNlnc endógeno. En tales casos, el promotor de ARNlnc impulsa la expresión de la secuencia del inserto de ácido nucleico. En un aspecto descrito en el presente documento, la secuencia del inserto de ácido nucleico es una secuencia de ácido nucleico indicadora.

En algunos casos, el inserto de ácido nucleico comprende un promotor. En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende un gen polinucleótido de interés unido operativamente a un promotor que impulsa la expresión del polinucleótido de interés. En un aspecto descrito en el presente documento, el polinucleótido de interés comprende una secuencia de ácido nucleico indicadora. En otro aspecto descrito en el presente documento, el polinucleótido de interés comprende una secuencia de ácido nucleico de un marcador de selección.

En un aspecto descrito en el presente documento, el promotor es un promotor constitutivamente activo.

En un aspecto descrito en el presente documento, el promotor es un promotor inducible. En un aspecto descrito en el presente documento, el promotor inducible es un promotor regulado químicamente. En un aspecto descrito en el presente documento, el promotor regulado químicamente es un promotor regulado por alcoholes. En un aspecto descrito en el presente documento, el promotor regulado por alcoholes es un promotor del gen de alcohol deshidrogenasa (alcA). En un aspecto descrito en el presente documento, el promotor regulado químicamente es un promotor regulado por tetraciclinas. En un aspecto descrito en el presente documento, el promotor regulado por tetraciclinas es un promotor sensible a tetraciclinas. En un aspecto descrito en el presente documento, el promotor regulado por tetraciclinas es una secuencia operadora de tetraciclinas (tetO). En un aspecto descrito en el presente documento, el promotor regulado por tetraciclinas es un promotor tet-On. En un aspecto descrito en el presente documento, el promotor regulado por tetraciclinas es un promotor tet-Off. En un aspecto descrito en el presente documento, el promotor regulado químicamente es un promotor regulado por esteroides. En un aspecto descrito en el presente documento, el promotor regulado por esteroides es un promotor de un receptor de glucocorticoides de rata. En un aspecto descrito en el presente documento, el promotor regulado por esteroides es un promotor de un receptor estrogénico. En un aspecto descrito en el presente documento, el promotor regulado por esteroides es un promotor de un receptor de ecdisonas. En un aspecto descrito en el presente documento, el promotor regulado químicamente es un promotor regulado por metales. En un aspecto descrito en el presente documento, el promotor regulado por metales es un promotor de metaloproteínas. En un aspecto descrito en el presente documento, el promotor inducible es un promotor regulado físicamente. En un aspecto descrito en el presente documento, el promotor regulado físicamente es un promotor regulado por temperatura. En un aspecto descrito en el presente documento, el promotor regulado por temperatura es un promotor de choque térmico. En un aspecto descrito en el presente documento, el promotor regulado físicamente es un promotor regulado por luz. En un aspecto descrito en el presente documento, el promotor regulado por luz es un promotor inducible por luz. En un aspecto descrito en el presente documento, el promotor regulado por luz es un promotor reprimible por luz.

En un aspecto descrito en el presente documento, el promotor es un promotor específico de tejido. En un aspecto descrito en el presente documento, el promotor es un promotor de neuronas. En un aspecto descrito en el presente documento, el promotor es un promotor específico de glía. En un aspecto descrito en el presente documento, el promotor es un promotor específico de células musculares. En un aspecto descrito en el presente documento, el promotor es un promotor específico de células cardíacas. En un aspecto descrito en el presente documento, el promotor es un promotor específico de células renales. En un aspecto descrito en el presente documento, el promotor es un promotor específico de células óseas. En un aspecto descrito en el presente documento, el promotor es un

promotor específico de células endoteliales. En un aspecto descrito en el presente documento, el promotor es un promotor específico de células inmunitarias. En un aspecto descrito en el presente documento, el promotor específico de células inmunitarias es un promotor de linfocitos B. En un aspecto descrito en el presente documento, el promotor específico de células inmunitarias es un promotor de linfocitos T.

5 En un aspecto descrito en el presente documento, el promotor es un promotor regulado por el desarrollo. En un aspecto descrito en el presente documento, el promotor regulado por el desarrollo es activo solo durante una etapa embrionaria del desarrollo. En un aspecto descrito en el presente documento, el promotor regulado por el desarrollo es activo solo en una célula adulta.

10 En aspectos específicos descritos en el presente documento, el promotor puede seleccionarse basándose en el tipo de célula. Así, los diferentes promotores encuentran uso en una célula eucariota, una célula de mamífero, una célula de mamífero no humano, una célula pluripotente, una célula pluripotente no humana, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula humana progenitora restringida por el desarrollo, una célula iPS humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster, un fibroblasto o una célula CHO.

20 En algunos aspectos descritos en el presente documento, el inserto de ácido nucleico comprende un ácido nucleico flanqueado con secuencias diana de recombinación específicas de sitio. Se reconoce todo el tiempo que el inserto de ácido nucleico puede estar flanqueado por tales secuencias diana de recombinación específica de sitio, cualquier región o polinucleótido individual de interés dentro del inserto de ácido nucleico puede también estar flanqueado por tales sitios. La recombinasa específica de sitio se puede introducir en la célula por cualquier medio, incluyendo la introducción del polipéptido de recombinasa en la célula o la introducción de un polinucleótido que codifica la recombinasa específica de sitio en la célula hospedadora. El polinucleótido que codifica la recombinasa específica de sitio se puede localizar dentro del inserto de ácido nucleico o dentro de un polinucleótido separado. La recombinasa específica de sitio puede estar unida operativamente a un promotor activo en la célula que incluye, por ejemplo, un promotor inducible, un promotor que es endógeno para la célula, un promotor que es heterólogo para la célula, un promotor específico de células, un promotor específicos de tejidos o un promotor específico de etapas de desarrollo. Las secuencias diana de recombinación específicas de sitio, que pueden flanquear el inserto de ácido nucleico o un polinucleótido de interés en el inserto de ácido nucleico pueden incluir, pero no se limitan a, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attp, att, FRT, rox o una combinación de los mismos.

30 En algunos aspectos descritos en el presente documento, los sitios de recombinación específica de sitios flanquean un polinucleótido que codifica un marcador de selección y/o un gen indicador contenido(s) dentro del inserto de ácido nucleico. En tales casos, tras la integración del inserto de ácido nucleico en el locus dirigido se pueden eliminar las secuencias entre los sitios de recombinación específica de sitios.

35 En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende un polinucleótido que codifica un marcador de selección. El marcador de selección puede estar contenido en un casete de selección. Tales marcadores de selección incluyen, pero no se limitan, a neomicina fosfotransferasa (neo^r), higromicina B fosfotransferasa (hyg^r), puromicina-N-acetiltransferasa (puro^r), blasticidina S desaminasa (bsr^r), xantina/guanina fosforribosil transferasa (gpt) o timidina quinasa del virus del herpes simple (HSV-TK) o una combinación de los mismos. En un aspecto descrito en el presente documento, el polinucleótido que codifica el marcador de selección está unido operativamente a un promotor activo en la célula. En un aspecto descrito en el presente documento, el polinucleótido que codifica el marcador de selección está flanqueado con secuencias diana de recombinación específicas de sitio.

40 El inserto de ácido nucleico puede comprender adicionalmente un gen indicador unido operativamente a un promotor, en donde el gen indicador codifica una proteína indicadora seleccionada del grupo que consiste en o que comprende β-galactosidasa (codificada por el gen lacZ), GFP, mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrina, Venus, YPet, proteína fluorescente amarilla potenciada (EYFP), Esmeralda, proteína fluorescente verde potenciada (EGFP), CyPet, proteína fluorescente cian (CFP), Cerulean, T-Zafiro, luciferasa, fosfatasa alcalina y/o una combinación de los mismos. Tales genes indicadores pueden estar unidos operativamente a un promotor activo en la célula. Tales promotores pueden ser un promotor inducible, un promotor que es endógeno para el gen indicador o la célula, un promotor que es heterólogo para el gen indicador o para la célula, un promotor específico de células, un promotor específicos de tejidos o un promotor específico de etapas de desarrollo.

50 En un aspecto descrito en el presente documento, la modificación genética comprende una delección de una secuencia no codificante de proteínas, pero no comprende una delección de una secuencia codificante de proteínas. En un aspecto descrito en el presente documento, la delección de la secuencia no codificante de proteínas comprende una delección de un locus de ARNlnc o porción del mismo. En un aspecto descrito en el presente documento, la delección de la secuencia no codificante de proteínas comprende una delección de un elemento regulador. En un aspecto descrito en el presente documento, la modificación genética comprende una delección de un elemento regulador. En un aspecto descrito en el presente documento, la modificación genética comprende una adición de un promotor o un elemento regulador. En un aspecto descrito en el presente documento, la modificación genética comprende un reemplazo de un promotor o un elemento regulador.

ii. Casetes de expresión

- 5 En el presente documento se proporcionan polinucleótidos o moléculas de ácidos nucleicos que comprenden los diversos componentes empleados en un sistema de integración genómico dirigido proporcionado en el presente documento para dirigirse a un locus de ARNInc (es decir, una cualquiera o cualquier combinación de agentes de nucleasas, sitios de reconocimiento, insertos de ácidos nucleicos, polinucleótidos de interés, secuencias indicadoras, vectores de direccionamiento, marcadores de selección y otros componentes).
- 10 Las expresiones "polinucleótido", "secuencia de polinucleótidos", "secuencia de ácidos nucleicos", y "fragmento de ácido nucleico" se usan indistintamente en el presente documento. Estas expresiones abarcan secuencias de nucleótidos y similares. Un polinucleótido puede ser un polímero de ARN o de ADN que es monocatenario o bicatenario, que opcionalmente contiene bases nucleotídicas sintéticas, no naturales o alteradas. Un polinucleótido en forma de un polímero de ADN puede estar comprendido de uno o más segmentos de ADNc, ADN genómico, ADN sintético o mezclas de los mismos. Los polinucleótidos pueden comprender desoxirribonucleótidos y ribonucleótidos, incluir tanto moléculas naturales como análogos sintéticos y cualquier combinación de estos. Los polinucleótidos proporcionados en el presente documento también abarcan todas las formas de secuencias incluyendo, pero sin limitarse a, formas monocatenarias, formas bicatenarias, horquillas, estructuras de tallo y bucle y similares.
- 20 En el presente documento se describen adicionalmente polinucleótidos que comprenden los diversos componentes del sistema de integración genómica dirigida para dirigirse a un locus de ARNInc. Las expresiones "polinucleótido recombinante" y "constructo de ADN recombinante" se usan indistintamente en el presente documento. Un constructo recombinante comprende una combinación artificial o heteróloga de secuencias de ácidos nucleicos, por ejemplo, secuencias reguladoras y codificantes que no se encuentran juntas en la naturaleza. En otros aspectos descritos en el presente documento, un constructo recombinante puede comprender secuencias reguladoras y secuencias codificantes que se derivan de diferentes fuentes o secuencias reguladoras y secuencias codificantes derivadas de la misma fuente, pero dispuestas de una manera diferente que la encontrada en la naturaleza. Dicho constructo puede utilizarse solo o se puede utilizar en conjunción con un vector. Si se utiliza un vector, a continuación, la elección del vector depende del método que se utilice para transformar las células hospedadoras, como se conoce bien por los expertos en la materia. Por ejemplo, puede usarse un vector plasmídico. Los elementos genéticos requeridos para transformar, seleccionar y propagar satisfactoriamente células hospedadoras que comprenden cualquiera de los fragmentos de ácidos nucleicos aislados proporcionados en el presente documento también se proporcionan. El cribado se puede conseguir mediante análisis de Southern de ADN, análisis de Northern de la expresión de ARNm, análisis de inmunotransferencia de la expresión de proteínas o análisis de fenotipos, entre otros.
- 35 En aspectos específicos descritos en el presente documento, uno o más de los componentes del sistema de integración genómica dirigida para dirigirse a un locus de ARNInc descrito en el presente documento puede proporcionarse en un casete de expresión para la expresión en una célula procariota, una célula eucariota, una bacteria, una levadura o una célula de mamífero u otro organismo o tipo de célula de interés. El casete puede incluir secuencias reguladoras 5' y 3' unidas operativamente con un polinucleótido proporcionado en el presente documento. "Unido operativamente" comprende una relación en donde los componentes unidos operativamente funcionan de la manera prevista. Por ejemplo, una unión operativa entre un polinucleótido de interés y una secuencia reguladora (es decir, un promotor) es un enlace funcional que permite la expresión del polinucleótido de interés. Los elementos unidos operativamente pueden ser contiguos o no contiguos. Cuando se usa para hacer referencia a la unión de dos regiones codificantes de proteínas, unido operativamente significa que las regiones codificantes están en el mismo marco de lectura. En otro caso, una secuencia de ácido nucleico que codifica una proteína puede unirse operativamente a secuencias reguladoras (por ejemplo, promotor, potenciador, secuencia silenciadora, etc.) con el fin de conservar la regulación de la transcripción adecuada. El casete puede contener adicionalmente al menos un polinucleótido adicional para cointroducirse en el organismo. Como alternativa, los polipéptidos de interés adicionales pueden proporcionarse en múltiples casetes de expresión. Tal casete de expresión se proporciona con una pluralidad de sitios de restricción y/o sitios de recombinación para que la inserción del polinucleótido esté bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes de marcadores de selección.
- 50 El casete de expresión puede incluir en la dirección de transcripción 5'-3', una región de inicio de la transcripción y de la traducción (es decir, un promotor), un polinucleótido recombinante proporcionado en el presente documento y una región de terminación de la transcripción y de la traducción (es decir, región de terminación) funcional en una célula de mamífero o en una célula hospedadora de interés. Las regiones reguladoras (es decir, promotores, regiones reguladoras transcripcionales, secuencia Kozak y regiones de terminación de la traducción) y/o un polinucleótido proporcionados en el presente documento pueden ser nativos/análogos a la célula hospedadora o entre sí. Como alternativa, las regiones reguladoras y/o un polinucleótido proporcionados en el presente documento pueden ser heterólogos de la célula hospedadora o entre sí. Por ejemplo, un promotor unido operativamente con un polinucleótido heterólogo es de una especie diferente de la especie de la que se derivó el polinucleótido o si es de la misma especie o una análoga, uno o ambos se modifican sustancialmente respecto a su forma y/o locus genómico original o el promotor no es el promotor nativo para el polinucleótido unido operativamente. Como alternativa, las regiones reguladoras y/o un polinucleótido recombinante proporcionados en el presente documento pueden ser sintéticos completamente.
- 65

La región de terminación puede ser nativa de la región de iniciación de la transcripción, puede ser nativa del polinucleótido recombinante unido operativamente, puede ser nativa de la célula hospedadora o puede derivarse de otra fuente (es decir, exógena o heteróloga) para el promotor, el polinucleótido recombinante, la célula hospedadora o cualquier combinación de los mismos.

En la preparación del casete de expresión, los diversos fragmentos de ADN pueden manipularse, para proporcionar las secuencias de ADN en la orientación apropiada. Con este fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN o pueden estar implicadas otras manipulaciones para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción o similares. Para este fin, pueden estar implicados mutagénesis *in vitro*, reparación de cebadores, restricción, hibridación, resustituciones, por ejemplo, transiciones y transversiones.

Se pueden utilizar varios promotores en los casetes de expresión proporcionados en el presente documento. Los promotores se pueden seleccionar basándose en el resultado deseado. Se reconoce que se pueden potenciar diferentes aplicaciones mediante el uso de promotores diferentes en los casetes de expresión para modular el momento, localización y/o nivel de expresión del polinucleótido de interés. Tales constructos de expresión pueden contener también, si se desea, una región reguladora de promotores (por ejemplo, una que confiere expresión inducible, constitutiva, regulada por el ambiente o por el desarrollo, o específica/selectiva de células o tejidos), un sitio de comienzo de iniciación de la transcripción, una secuencia consenso Kozak, un sitio de unión al ribosoma, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación.

El casete de expresión que contiene los polinucleótidos proporcionados en el presente documento también pueden comprender un gen marcador de selección para la selección de células transformadas. Se utilizan genes marcadores seleccionables para la selección de células o tejidos transformados.

Cuando sea adecuado, las secuencias empleadas en los métodos y composiciones (es decir, el polinucleótido de interés, el agente nucleasa, etc.) puede optimizarse para su expresión incrementada en la célula. Es decir, los genes se pueden sintetizar utilizando codones preferidos en una célula de interés dada, incluyendo, por ejemplo, codones preferidos de mamíferos, codones preferidos humanos, codones preferidos de roedores, codones preferidos de ratones, codones preferidos de ratas, codones preferidos de hámster, etc. para una expresión mejorada.

Los diversos métodos y composiciones proporcionadas en el presente documento pueden emplear marcadores de selección. Se pueden utilizar diversos marcadores de selección en los métodos y composiciones desvelados en el presente documento. Tales marcadores de selección pueden, por ejemplo, conferir resistencia a un antibiótico tal como el G418, higromicina, blastomicina, neomicina o puromicina. Tales marcadores de selección incluyen neomicina fosfotransferasa (*neo^r*), higromicina B fosfotransferasa (*hyg^r*), puromicina-N-acetiltransferasa (*puro^r*) y blastomicina S desaminasa (*bsr^r*). En todavía otros aspectos descritos en el presente documento, el marcador de selección está unido operativamente a un promotor inducible y la expresión del marcador de selección es tóxica para la célula. Los ejemplos no limitantes de tales marcadores de selección incluyen xantina/guanina fosforribosil transferasa (*gpt*), haixipantantina-guanina fosforribosiltransferasa (HGPRT) o timidina quinasa del virus del herpes simple (HSV-TK). El polinucleótido que codifica los marcadores de selección está unido operativamente a un promotor activo en la célula.

iii. Vectores de direccionamiento

Los vectores de direccionamiento se emplean para introducir el inserto de ácido nucleico en el locus de ARN_{lnc} de interés del ácido nucleico eucariota, no humano, de mamífero, de mamífero no humano, ser humano, de roedor, de ratón, de rata o de hámster. El vector de direccionamiento comprende el inserto de ácido nucleico y adicionalmente comprende un brazo de homología 5' y 3', que flanquea el inserto de ácido nucleico. Los brazos de homología, que flanquean el inserto de ácido nucleico, se corresponden con regiones dentro del locus de ARN_{lnc} diana del ácido nucleico eucariota, no humano, de mamífero, de mamífero no humano, ser humano, de roedor, de ratón, de rata o de hámster. Para facilitar la referencia, las regiones genómicas afines correspondientes dentro del locus genómico dirigido se denominan "sitios diana". Por ejemplo, un vector de direccionamiento puede comprender un primer inserto de ácido nucleico flanqueado por un primer y un segundo brazo de homología complementarios de un primer y un segundo sitio diana. Como tal, vector de direccionamiento ayuda de ese modo a la integración del inserto de ácido nucleico en el ácido nucleico de locus diana a través de un evento de recombinación homóloga que se produce entre los brazos de homología y los sitios diana complementarios dentro del genoma de la célula.

En un aspecto descrito en el presente documento, el locus diana del ácido nucleico eucariota, de mamífero, de mamífero no humano, ser humano, de roedor, de ratón o hámster, comprende una primera secuencia de ácido nucleico que es complementaria del brazo de homología 5' y una segunda secuencia de ácido nucleico que es complementaria del brazo de homología 3'. En un aspecto descrito en el presente documento, la primera y la segunda secuencia de ácido nucleico están separadas por al menos 5kb. En otro aspecto descrito en el presente documento, la primera y la segunda secuencia de ácido nucleico están separadas por al menos 1kb pero menos de 50kb. En un aspecto descrito en el presente documento, la primera y la segunda secuencia de ácido nucleico están separadas por al menos 2kb. En un aspecto descrito en el presente documento, la primera y la segunda secuencia de ácido nucleico están

separadas por al menos 3kb, al menos 4kb, al menos 5kb, al menos 6kb, al menos 7b, al menos 8kb, al menos 9kb, al menos 10kb, al menos 15kb, al menos 20kb, al menos 30kb, al menos 40kb o al menos 50kb. En todavía aspectos adicionales descritos en el presente documento, la primera y la segunda secuencia de ácido nucleico están separadas por al menos 1kb pero menos de 2kb, al menos 2kb pero menos de 3kb, al menos 4kb pero menos de 5kb, al menos 5kb pero menos de 6kb, al menos 6kb pero menos de 7kb, al menos 7kb pero menos de 8kb, al menos aproximadamente 8kb pero menos de 9kb, al menos 9kb pero menos de 10kb o al menos 10kb pero menos de 15kb, al menos aproximadamente 15kb pero menos de aproximadamente 20kb, al menos aproximadamente 20kb pero menos de aproximadamente 30kb o al menos alrededor de 40kb pero menos de aproximadamente 50kb.

10 Un brazo de homología del vector de direccionamiento puede tener cualquier longitud que sea suficiente para promover un evento de recombinación homólogo con un sitio diana correspondiente, que incluye por ejemplo, al menos 5-10kb, 5-15kb, 10-20kb, 20-30kb, 30-40kb, 40-50kb, 50-60kb, 60-70kb, 70-80kb, 80-90kb, 90-100kb, 100-110kb, 110-120kb, 120-130kb, 130-140kb, 140-150kb, 150-160kb, 160-170kb, 170-180kb, 180-190kb, 190-200kb de longitud o más. Como se esboza en más detalle más adelante, los vectores de direccionamiento grandes pueden emplear brazos de direccionamiento de mayor longitud. En un aspecto específico descrito en el presente documento, la suma total del brazo de homología 5' y el brazo de homología 3' es de al menos 10 kb o la suma total del brazo de homología 5' y el brazo de homología 3' es de al menos aproximadamente 16kb a aproximadamente 100 kb o de aproximadamente 30kb a aproximadamente 100kb. En otros aspectos descritos en el presente documento, el tamaño de la suma total del total de los brazos de homología 5' y 3' de LTVEC es de aproximadamente 10kb a aproximadamente 150kb, de aproximadamente 10kb a aproximadamente 100kb, de aproximadamente 10kb a aproximadamente 75kb, de aproximadamente 20kb a aproximadamente 150kb, de aproximadamente 20kb a aproximadamente 100kb, de aproximadamente 20kb a aproximadamente 75kb, de aproximadamente 30kb a aproximadamente 150kb, de aproximadamente 30kb a aproximadamente 100kb, de aproximadamente 30kb a aproximadamente 75kb, de aproximadamente 40kb a aproximadamente 150kb, de aproximadamente 40kb a aproximadamente 100kb, de aproximadamente 40kb a aproximadamente 75kb, de aproximadamente 50kb a aproximadamente 150kb, de aproximadamente 50kb a aproximadamente 100kb o de aproximadamente 50kb a aproximadamente 75kb, de aproximadamente 10kb a aproximadamente 30kb, aproximadamente 20kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 120kb o de aproximadamente 120kb a aproximadamente 150kb. En un aspecto descrito en el presente documento, el tamaño de la delección es el mismo o similar al tamaño de la suma total de los brazos de homología 5' y 3' del LTVEC.

Un brazo de homología y un sitio diana (es decir, una región genómica afín) se "complementan" o son "complementarios" entre sí cuando las dos regiones comparten un nivel suficiente de identidad de secuencia entre sí para actuar como sustratos para una reacción de recombinación homóloga. Por "homología" se entiende las secuencias de ADN que son idénticas o que comparten la identidad de secuencia con una secuencia correspondiente o "complementaria". La identidad de secuencia entre un sitio diana dado y el brazo de homología correspondiente encontrado en el vector de direccionamiento puede ser de cualquier grado de identidad de secuencia que permita que suceda una recombinación homóloga. Por ejemplo, la cantidad de identidad de secuencia compartida por el brazo de homología del vector de direccionamiento (o un fragmento del mismo) y el sitio diana (o un fragmento del mismo) puede ser de al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia, de manera que las secuencias se someten a recombinación homóloga. Por otra parte, una región complementaria de homología entre el brazo de homología y el sitio diana complementario puede ser de cualquier longitud que sea suficiente para promover la recombinación homóloga en el sitio de reconocimiento escindido. Por ejemplo, un brazo de homología dado y/o un sitio diana complementario puede(n) comprender regiones complementarias de homología que son al menos de 5-10kb, 5-15kb, 10-20kb, 20-30kb, 30-40kb, 40-50kb, 50-60kb, 60-70kb, 70-80kb, 80-90kb, 90-100kb, 100-110kb, 110-120kb, 120-130kb, 130-140kb, 140-150kb, 150-160kb, 160-170kb, 170-180kb, 180-190kb, 190-200kb, 200kb a 300kb de longitud o más (tal como se describió en los vectores LTVEC en otra parte del presente documento) de manera que el brazo de homología tiene homología suficiente para someterse a recombinación homóloga con los sitios diana correspondientes dentro del genoma de la célula. Para facilitar la referencia, los brazos de homología se denominan en el presente documento como un brazo de homología 5' y un brazo de homología 3'. Esta terminología se refiere a la posición relativa de los brazos de homología hacia el inserto de ácido nucleico dentro del vector de direccionamiento.

55 Los brazos de homología del vector de direccionamiento están por tanto diseñados para ser complementarios de un sitio diana con el locus dirigido. Por tanto, los brazos de homología pueden ser complementarios de un locus que es nativo para la célula o alternativamente, pueden ser complementarios de una región de un segmento heterólogo o exógeno de ADN que se integró en el genoma de la célula, incluyendo, pero sin limitarse a, transgenes, casetes de expresión o regiones heterólogas o exógenas de ADN genómico. Como alternativa, los brazos de homología del vector de direccionamiento pueden ser complementarios de una región de un cromosoma artificial humano o de cualquier otra región genómica diseñada contenida en una célula hospedadora adecuada. Más aún, los brazos de homología del vector de direccionamiento pueden ser complementarios de o derivados de una región de una biblioteca BAC, una biblioteca de cósmidos o una biblioteca de fagos PI. Por tanto, en aspectos específicos descritos en el presente documento, los brazos de homología del vector de direccionamiento son complementarios de un locus genómico eucariota, no humano, de mamífero, de mamífero no humano, ser humano, de roedor, de ratón o de rata que es nativo,

heterólogo o exógeno para una célula dada. En un aspecto descrito en el presente documento, los brazos de homología se derivan de ADN sintético.

5 El vector de direccionamiento (tal como un vector de direccionamiento grande) también puede comprender un casete de selección o un gen indicador como se analiza en otra parte en el presente documento. El casete de selección puede comprender una secuencia de ácido nucleico que codifica un marcador de selección, en donde la secuencia de ácido nucleico está unida operativamente a un promotor como se analiza en otra parte en el presente documento. El marcador de selección y/o el gen indicador del vector de direccionamiento puede(n) estar flanqueado por los brazos de homología 5' y 3' o encontrarse 5' o 3' hacia los brazos de homología.

10 En un aspecto descrito en el presente documento, un vector de direccionamiento comprende un inserto de ácido nucleico que comprende una primera secuencia de nucleótidos que codifica un indicador. En algunos casos, tras la recombinación homóloga con el locus de ARNinc de interés, la primera secuencia de nucleótidos que codifica el indicador está ligada operativamente a un promotor endógeno que impulsa la expresión de un ARNinc en el locus de ARNinc. En un aspecto adicional descrito en el presente documento, la secuencia del inserto de ácido nucleico del vector de direccionamiento comprende una secuencia consenso Kozak. En tales casos donde el inserto de ácido nucleico comprende un indicador, la secuencia consenso Kozak puede estar unida operativamente a la secuencia de ácido nucleico que codifica el indicador.

15 En otro aspecto descrito en el presente documento, el inserto de ácido nucleico del vector de direccionamiento comprende una segunda secuencia de nucleótidos que codifica un marcador seleccionable. En algunos casos, el segundo ácido nucleico está ligado operativamente a un promotor.

20 En un aspecto descrito en el presente documento, la primera y/o segunda secuencia de nucleótidos del inserto de ácido nucleico comprende(n) una secuencia consenso Kozak.

25 En un aspecto descrito en el presente documento, el vector de direccionamiento (tal como un vector de direccionamiento grande) comprende un gen indicador y/o un gen marcador seleccionable unido operativamente a un promotor como se describe en otra parte en el presente documento. Tales genes indicadores y/o genes de marcadores seleccionables pueden estar unidos operativamente a un promotor activo en la célula como se describe en otra parte en el presente documento.

30 En un aspecto descrito en el presente documento, el vector de direccionamiento comprende un gen de recombinasa específico de sitio. En un aspecto descrito en el presente documento, el gen de recombinasa específico de sitio codifica una recombinasa Cre. En un aspecto descrito en el presente documento, el gen de la recombinasa Cre es Crei, en donde dos exones que codifican la recombinasa Cre están separados por un intrón para evitar su expresión en una célula procarionota. En un aspecto descrito en el presente documento, el gen de recombinasa específico de sitio codifica una recombinasa Dre.

35 En un aspecto descrito en el presente documento, el gen de recombinasa Cre comprende adicionalmente una señal de localización nuclear para facilitar la localización de Cre (o cualquier agente de recombinasa o nucleasa) hacia el núcleo (por ejemplo, el gen es un gen NL-Cre). En un aspecto específico descrito en el presente documento, el gen de recombinasa Cre comprende adicionalmente una señal de localización nuclear y un intrón (por ejemplo, NL-Crei).

40 En diversos aspectos descritos en el presente documento, un promotor adecuado para la expresión de la recombinasa Cre o Crei analizada anteriormente se selecciona de o comprende un Prm1, Blimp1, Gata6, Gata4, Igf2, Lhx2, Lhx5 y/o Pax3. En un aspecto específico descrito en el presente documento, el promotor es el promotor Gata6 o Gata4. Los diversos promotores pueden ser de cualquier organismo, que incluye por ejemplo, un roedor tal como un ratón o una rata, un eucariota, un mamífero no humano, un mamífero, un ser humano o un hámster. En otro aspecto específico descrito en el presente documento, el promotor es un promotor Prm1 de ratón. En otro aspecto específico descrito en el presente documento, el promotor es un promotor Blimp1 o un fragmento del mismo, por ejemplo, un fragmento de 1 kb o 2 kb de un promotor Blimp1. Véanse, por ejemplo, la patente de Estados Unidos N.º 8.697.851 y la Publicación de Solicitud de Estados Unidos 2013-0312129.

45 En un aspecto descrito en el presente documento, el inserto de ácido nucleico comprende una secuencia nucleotídica flanqueada por dos sitios de recombinación específica de sitio. Entre los ejemplos de sitios de recombinación específica de sitio se incluyen, pero no se limitan a, loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox y una combinación de los mismos.

50 *iv. Vectores de direccionamiento grandes*

55 El término "vector de direccionamiento grande" o "LTVEC" incluye vectores de direccionamiento grandes que comprenden brazos de homología que se corresponden con y se derivan de secuencias de ácido nucleico más grandes que las utilizadas típicamente por otros enfoques previstos para realizar un direccionamiento homólogo en células y/o que comprenden insertos de polinucleótidos que comprenden secuencias de ácido nucleico más grandes que las

utilizadas típicamente por otros enfoques destinados a realizar un direccionamiento de recombinación homóloga en células. En aspectos específicos descritos en el presente documento, los brazos de homología y/o el inserto de polinucleótidos de los LTVEC comprenden una secuencia genómica de una célula eucariota. El tamaño del LTVEC es demasiado grande para permitir el cribado de eventos de direccionamiento mediante ensayos convencionales, por ejemplo, transferencia de Southern y PCR de largo alcance (por ejemplo, 1kb-5kb). Entre los ejemplos de los LTVEC, se incluyen, pero no se limitan a, vectores derivados de un cromosoma artificial bacteriano (BAC), un cromosoma artificial humano o un cromosoma artificial de levadura (YAC). Se describen ejemplos no limitativos de LTVEC y métodos para hacerlos, por ejemplo, en las Patentes de Estados Unidos N.º 6.586.251, 6.596.541, 7.105.348 y WO 2002/036789 (PCT/US01/45375).

El LTVEC puede ser de cualquier longitud, incluyendo, pero sin limitarse a, al menos aproximadamente 10kb, aproximadamente 15kb, aproximadamente 20kb, aproximadamente 30kb, aproximadamente 40kb, aproximadamente 50kb, aproximadamente 60kb, aproximadamente 70kb, aproximadamente 80kb, aproximadamente 90kb, aproximadamente 100kb, aproximadamente 150kb, aproximadamente 200kb, de aproximadamente 10kb a aproximadamente 15kb, de aproximadamente 15 kb a aproximadamente 20kb, aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 300kb, de aproximadamente 50kb a aproximadamente 75kb, de aproximadamente 75kb a aproximadamente 100kb, de aproximadamente 100kb a 125kb, de aproximadamente 125kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 175kb, aproximadamente 175kb a aproximadamente 200kb, de aproximadamente 200kb a aproximadamente 225kb, de aproximadamente 225kb a aproximadamente 250kb, de aproximadamente 250kb a aproximadamente 275kb o de aproximadamente 275kb a aproximadamente 300kb.

En un aspecto descrito en el presente documento, los brazos de homología del LTVEC se derivan de una biblioteca BAC, una biblioteca de cósmidos o una biblioteca de fagos PI. En otros aspectos descritos en el presente documento, los brazos de homología se derivan del locus genómico de ARNInc dirigido de la célula y en algunos casos, el locus genómico diana para el que el LTVEC está diseñado a dirigirse no es dirigible utilizando un método convencional. En todavía otros aspectos descritos en el presente documento, los brazos de homología se derivan de ADN sintético.

En un aspecto descrito en el presente documento, una suma total del brazo de homología en dirección 5' y del brazo de homología en dirección 3' en el LTVEC es de al menos 10kb. En otro aspecto descrito en el presente documento, el brazo de homología en dirección 5' varía de aproximadamente 5kb a aproximadamente 100kb. En un aspecto descrito en el presente documento, el brazo de homología en dirección 3' varía de aproximadamente 5kb a aproximadamente 100kb. En otros aspectos descritos en el presente documento, la suma total de los brazos de homología en dirección 5' y en dirección 3' varía de aproximadamente 5kb a aproximadamente 10kb, de aproximadamente 10kb a aproximadamente 20kb, de aproximadamente 20kb a aproximadamente 30kb, de aproximadamente 30kb a aproximadamente 40kb, de aproximadamente 40kb a aproximadamente 50kb, de aproximadamente 50kb a aproximadamente 60kb, de aproximadamente 60kb a aproximadamente 70kb, de aproximadamente 70kb a aproximadamente 80kb, de aproximadamente 80kb a aproximadamente 90kb, de aproximadamente 90kb a aproximadamente 100kb, de aproximadamente 100kb a aproximadamente 110kb, de aproximadamente 110kb a aproximadamente 120kb, de aproximadamente 120kb a aproximadamente 130kb, de aproximadamente 130kb a aproximadamente 140kb, de aproximadamente 140kb a aproximadamente 150kb, de aproximadamente 150kb a aproximadamente 160kb, de aproximadamente 160kb a aproximadamente 170kb, de aproximadamente 170kb a aproximadamente 180kb, de aproximadamente 180kb a aproximadamente 190kb o de aproximadamente 190kb a aproximadamente 200kb. En un aspecto descrito en el presente documento, el tamaño de la delección es el mismo o similar al tamaño de la suma total de los brazos de homología 5' y 3' del LTVEC.

En un aspecto descrito en el presente documento, el LTVEC comprende un casete de selección o un gen indicador como se ha descrito en otra parte en el presente documento.

III. Métodos de introducción de secuencias y generación de animales transgénicos

Como se ha destacado anteriormente, se proporcionan métodos y composiciones en el presente documento para la modificación genética dirigida de uno o más loci de ARNInc. Se reconoce adicionalmente que puede hacerse una modificación genética dirigida adicional. Tales sistemas que permiten estas modificaciones genéticas dirigidas pueden emplear diversos componentes y para facilitar la referencia, en el presente documento la expresión sistema de integración genómica dirigida de forma genérica incluye todos los componentes requeridos para un evento de integración (es decir, los diversos agentes de nucleasas, sitios de reconocimiento, insertos de polinucleótidos de ADN, vectores de direccionamiento, locus genómico diana y polinucleótidos de interés).

Los métodos proporcionados en el presente documento comprenden introducir en una célula uno o más constructos de polinucleótidos o de polipéptidos que comprenden los diferentes componentes del sistema de integración genómica dirigida. "Introducción" significa presentar a la célula la secuencia (polipéptido o polinucleótido) de tal manera que la secuencia tiene acceso al interior de la célula. Los métodos proporcionados en el presente documento no dependen de un método particular para introducir cualquier componente del sistema de integración genómica dirigida en la célula, solo que el polinucleótido tiene acceso al interior de al menos una célula. Se conocen métodos en la técnica métodos para introducir polinucleótidos en diversos tipos celulares incluyendo, pero no se limitan a, métodos de transfección

estables, métodos de transfección transitoria y métodos mediados por virus.

En algún aspecto descrito en el presente documento, las células empleadas en los métodos y composiciones tienen un constructo de ADN incorporado establemente en su genoma. "Incorporado establemente" o "introducido de establemente" significa la introducción de un polinucleótido en la célula de manera que la secuencia de nucleótidos se integra en el genoma de la célula y puede heredarse por la progenie de la misma. Se puede usar cualquier protocolo para la incorporación estable de los constructos de ADN o de los diversos componentes del sistema de integración genómica dirigida.

Los protocolos de transfección, así como los protocolos para introducir secuencias de polipéptidos o de polinucleótidos en las células, pueden variar. Entre los métodos no limitantes de transfección se incluyen los métodos de transfección basados en productos químicos, se incluye el uso de liposomas; nanopartículas; fosfato cálcico (Graham *et al.* (1973). *Virology* 52 (2): 456-67, Bacchetti *et al.* (1977) *Proc Natl Acad Sci USA* 74 (4): 1590-4 y Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. Nueva York: W. H. Freeman and Company. págs. 96-97); dendrímeros; o polímeros catiónicos tales como DEAE-dextrano o polietilénimina. Los métodos no químicos incluyen la electroporación; sonoporación; y transfección óptica. La transfección basada en partículas incluye el uso de una pistola de genes, transfección asistida por imán (Bertram, J. (2006) *Current Pharmaceutical Biotechnology* 7, 277-28). También pueden utilizarse métodos víricos para la transfección.

Se pueden generar animales no humanos, empleando los diversos métodos develados en el presente documento. Tales métodos comprenden (1) integrar uno o más polinucleótidos de interés, en el locus genómico de ARNinc diana de interés de una célula pluripotente del animal no humano para generar una célula pluripotente genéticamente modificada que comprende el inserto de polinucleótidos en el locus genómico de ARNinc dirigido empleando los métodos develados en el presente documento; (2) seleccionar la célula pluripotente modificada genéticamente que tiene uno o más polinucleótidos de interés en el locus genómico de ARNinc diana; (3) introducir la célula pluripotente modificada genéticamente en un embrión hospedador del animal no humano, por ejemplo, en una etapa premórula; y (4) implantar el embrión hospedador que comprende la célula pluripotente genéticamente modificada en una madre subrogada para generar una generación F0 derivada de la célula pluripotente modificada genéticamente. Se pueden emplear métodos similares para dirigirse a un locus cromosómico diana. El animal no humano puede ser un mamífero no humano, un roedor, un ratón, una rata, un hámster, un mono, un mamífero agrícola o un mamífero doméstico o un pez o un ave.

La célula pluripotente puede ser una célula ES humana, una célula ES no humana, una célula ES de roedor, una célula ES de ratón, una célula ES de rata, una célula ES de hámster, una célula ES de mono, una célula ES de mamífero agrícola o una célula ES de mamífero domesticado. En otros aspectos descritos en el presente documento, la célula pluripotente es una célula no humana, una célula de mamífero, célula humana, una célula de mamífero no humano, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula humana progenitora restringida por el desarrollo, una célula iPS humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster. En un aspecto descrito en el presente documento, la modificación genética dirigida da como resultado la pérdida de función del ARNinc.

Se pueden utilizar también técnicas de transferencia nuclear para generar animales no humanos. En resumen, los métodos para la transferencia nuclear incluyen las etapas de: (1) enuclear un ovocito; (2) aislar una célula o núcleo donante para combinar con el ovocito enucleado; (3) insertar la célula o núcleo en el ovocito enucleado para formar una célula reconstituida; (4) implantar la célula reconstituida en el útero de un animal para formar un embrión; y (5) permitir que el embrión se desarrolle. En tales métodos, los ovocitos habitualmente se obtienen de animales muertos, aunque también pueden aislarse de oviductos y/o ovarios de animales vivos. Los ovocitos pueden madurar en diversos medios conocidos por los expertos en la técnica antes de la enucleación. La enucleación del ovocito se puede realizar de varias maneras bien conocidas por los expertos en la materia. La inserción de la célula o núcleo donante en el ovocito enucleado para formar una célula reconstituida se realiza normalmente mediante microinyección de una célula donante bajo la zona pelúcida antes de la fusión. La fusión se puede inducir mediante la aplicación de un pulso eléctrico de CC a través del plano de contacto/fusión (electrofusión), mediante la exposición de las células a sustancias químicas que promueven la fusión, tales como el polietilenglicol o mediante un virus inactivado, como el virus Sendai. Una célula reconstituida se activa típicamente por medios eléctricos y/o no eléctricos, es decir, antes, durante y/o después de la fusión del donante nuclear y el ovocito receptor. Los métodos de activación incluyen pulsos eléctricos, choque inducido químicamente, penetración mediante esperma, niveles incrementados de cationes divalentes en el ovocito y reducción de la fosforilación de proteínas celulares (como mediante los inhibidores de la quinasa) en el ovocito. Las células reconstituidas activadas o embriones, se cultivan típicamente en un medio bien conocido por los expertos en la materia y a continuación se transfieren al útero de un animal. Véanse, por ejemplo, los documentos US20080092249, WO/1999/005266A2, US20040177390, WO/2008/017234A1 y la Patente de Estados Unidos N°. 7.612.250.

Se proporcionan otros métodos para hacer que un animal no humano comprenda en su línea germinal una o más modificaciones genéticas, como se ha descrito en el presente documento, que comprenden: (a) modificar un locus de ARNinc genómico dirigido de un animal no humano en una célula procariota que emplea los diversos métodos descritos en el presente documento; (b) seleccionar una célula procariota modificada que comprende la modificación

genética en el locus genómico dirigido; (c) aislar el vector de direccionamiento modificado genéticamente del genoma de la célula procariota modificada; (d) introducir el vector de direccionamiento modificado genéticamente en una célula pluripotente del animal no humano para generar una célula pluripotente modificada genéticamente que comprende el inserto de ácido nucleico en el locus genómico de ARNinc dirigido; (e) seleccionar la célula pluripotente modificada genéticamente; (f) introducir la célula pluripotente modificada genéticamente en un embrión hospedador del animal no humano en una etapa premórula; y (g) implantar el embrión hospedador que comprende la célula pluripotente genéticamente modificada en una madre subrogada para generar una generación F0 derivada de la célula pluripotente modificada genéticamente. En tales métodos los vectores de direccionamiento pueden comprender un vector de direccionamiento grande. El animal no humano puede ser un mamífero no humano, un roedor, un ratón, una rata, un hámster, un mono, un mamífero agrícola o un mamífero doméstico. La célula pluripotente puede ser una célula ES humana, una célula ES no humana, una célula ES de roedor, una célula ES de ratón, una célula ES de rata, una célula ES de hámster, una célula ES de mono, una célula ES de mamífero agrícola o una célula ES de mamífero doméstico. En otros aspectos descritos en el presente documento, la célula pluripotente es una célula no humana, célula de mamífero, célula humana, una célula de mamífero no humano, una célula pluripotente humana, una célula ES humana, una célula madre adulta humana, una célula humana progenitora restringida por el desarrollo, una célula IPS humana, una célula humana, una célula de roedor, una célula de rata, una célula de ratón, una célula de hámster. En un aspecto descrito en el presente documento, la modificación genética dirigida da como resultado la pérdida de función del ARNinc.

En métodos adicionales, la etapa de aislamiento (c) comprende adicionalmente (c1) linealizar el vector de direccionamiento modificado genéticamente (es decir, el LTVEC modificado genéticamente). En todavía aspectos adicionales descritos en el presente documento, la etapa de introducción (d) comprende adicionalmente (d1) introducir un agente de nucleasa en la célula pluripotente para facilitar la recombinación homóloga. En un aspecto descrito en el presente documento, las etapas de selección (b) y/o (e) se llevan a cabo mediante la aplicación de un agente seleccionable como se describe en el presente documento a la célula procariota o a la célula pluripotente. En un aspecto descrito en el presente documento, las etapas de selección (b) y/o (e) se llevan a cabo a través de un ensayo de modificación de alelos (MOA; del inglés, modification of allele assay) como se ha descrito en el presente documento.

En algunos aspectos descritos en el presente documento, pueden llevarse a cabo diversas modificaciones genéticas de los loci genómicos diana descritos en el presente documento mediante una serie de reacciones de recombinación homóloga (BHR; del inglés, homologous recombination reactions) en células bacterianas utilizando un LTVEC derivado de ADN del Cromosoma Artificial Bacteriano (BAC; del inglés, Bacterial Artificial Chromosome) utilizando la tecnología de ingeniería genética VELOCIGENE® (véanse, por ejemplo, la patente de Estados Unidos N.º 6.586.251 y Valenzuela, D. M. *et al.* (2003), *Nature Biotechnology* 21(6): 652-659).

En algunos aspectos descritos en el presente documento, las células pluripotentes y/o totipotentes dirigidas de ARNinc que comprenden diversas modificaciones genéticas como se describen en el presente documento se utilizan como insertos de células donantes y se introducen en un embrión en etapa premórula de un organismo correspondiente, por ejemplo, un embrión de ratón en etapa de 8 células, a través del método VELOCIMOUSE® (véanse, por ejemplo, los documentos US 7.576.259, US 7.659.442, US 7.294.754 y US 2008-0078000 A1). El embrión de animal no humano que comprende las células pluripotentes y/o totipotentes modificadas genéticamente se incuba hasta la etapa de blastocisto y a continuación se implanta en una madre subrogada para producir una generación F0. En algunos aspectos descritos en el presente documento, las células ES de mamífero dirigidas que comprenden diversas modificaciones genéticas como se describe en el presente documento se introducen en un embrión en etapa de blastocisto. Los animales no humanos que portan el locus genómico modificado genéticamente (es decir, un locus de ARNinc) se pueden identificar a través de un ensayo de modificación de alelos (MOA) como se ha descrito en el presente documento. El animal no humano de generación F0 resultante derivado de las células pluripotentes y/o totipotentes modificadas genéticamente se cruza con un animal no humano de tipo silvestre para obtener descendencia de generación F1. Tras el genotipado con cebadores y/o sondas específicas, los animales no humanos F1 que son heterocigotos para el locus genómico modificado genéticamente se cruzan entre sí para producir descendencia de animales no humanos de generación F2 que son homocigotos para el locus genómico modificado genéticamente.

En un aspecto descrito en el presente documento, se proporciona un método para hacer un animal no humano que comprende una modificación genética en al menos un locus de ARNinc. Tal método comprendiendo: (a) poner en contacto una célula pluripotente con un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3'; en donde el constructo de direccionamiento sufre recombinación homóloga con el locus de ARNinc en un genoma de la célula para formar una célula pluripotente modificada; (b) introducir la célula pluripotente modificada en un embrión hospedador; y (c) gestar el embrión de hospedador en una madre subrogada, en donde la madre subrogada produce progenie que comprende un locus de ARNinc modificado, en donde dicha modificación genética da como resultado una pérdida de función de al menos un ARNinc.

IV. Células

Los diversos métodos descritos en el presente documento emplean un sistema de direccionamiento de locus genómico para modificar un locus de ARNinc en una célula. Tales células incluyen células procariotas tales como células

bacterianas que incluyen *E. coli* o células eucariotas tales como levaduras, células de insectos, de anfibio, de planta o de mamífero, que incluyen, pero sin limitarse a una célula de ratón, una célula de rata, una célula de hámster, una célula de conejo, una célula de cerdo, una célula bovina, una célula de ciervo, una célula de oveja, una célula de cabra, una célula de pollo, una célula de gato, una célula de perro, una célula de hurón, una célula de primate (por ejemplo, mono tití, macaco rhesus) y similares y células de mamíferos domesticados o células de mamíferos agrícolas. Algunas células no son humanas, en particular células de mamífero no humano. En algunos aspectos descritos en el presente documento, para aquellos mamíferos para los que las células pluripotentes modificables genéticamente adecuadas no estén fácilmente disponibles, se emplearon otros métodos para reprogramar las células somáticas en células pluripotentes, por ejemplo, a través de la introducción en células somáticas de una combinación de factores de inducción de pluripotencia, que incluyen, pero sin limitarse a, Oct3/4, Sox2, KLF4, Myc, Nanog, LIN28 y Glis1. En tales métodos, la célula también puede ser una célula de mamífero, célula humana, una célula de mamífero no humano, una célula no humana, una célula de un roedor, una rata, un ratón, un hámster, una celular de fibroblastos o cualquier otra célula hospedadora. En otros aspectos descritos en el presente documento, la célula es una célula pluripotente, una célula pluripotente inducida humana (iPS), una célula madre embrionaria (ES) no humana. Tales células incluyen células pluripotentes, que incluyen, por ejemplo, células pluripotentes inducidas humanas (iPS), células iPS humanas, células madre embrionarias (ES) de ratón, células madre embrionarias (ES) de rata, células madre embrionarias (ES) humanas o células progenitoras humanas restringidas por el desarrollo, una célula madre embrionaria (ES) de roedor, una célula madre embrionaria (ES) de ratón o una célula madre embrionaria (ES) de rata.

20 Ejemplos

Ejemplo 1: Construcción de vectores de direccionamiento

Se emplearon métodos VelociGene®, como se ha descrito anteriormente, lo que permitió la generación rápido y de alto rendimiento de mutaciones génicas en ratones (Valenzuela, D.M., *et al.* (2003b), *Nat Biotechnol* 21:652-659). En resumen, se generaron vectores de direccionamiento grandes (LTVEC) BacVec utilizando clones BAC bMQ (129S7/SvEv Brd-Hprt b-m2) o RP23 BAC library (Adams, D.J., *et al.* (2005), *Genomics* 86:753-758). El casete de indicador/selección *lacZ/neo^r* (Fig. 1) fue idéntico al casete ZEN-Ub1 utilizado para el NIH KOMP (secuencia disponible en www.velocigene.com/komp/detail/10020) excepto en que el extremo amino terminal de la secuencia codificante de β-galactosidasa en la parte *lacZ* se modificó para incluir un codón de iniciación ATG y una secuencia consenso Kozak (Kozak, M. (1987) *Nucleic acids research* 15: 8125-8148)).

Ejemplo 2: Direccionamiento de células ES

Se introdujeron LTVEC en células ES híbridas F1 VGF1 (129S6SvEvTac/C57BL6NTac) (Poueymirou *et al.* (2007); Valenzuela *et al.* (2003a)) con un dispositivo de electroporación multipocillo (Harvard Apparatus, Boston, Massachusetts) en tampón de electroporación (Millipore) 3,3 x 10⁶ células, 0,67 µg de ADN en un volumen de 0,125 ml seguido de cultivo en placas gelatinizadas de 15 cm. El medio de selección que contiene G418 se añadió 48 horas después de la electroporación y se cambió diariamente a partir de entonces. Las colonias resistentes a fármacos se extrajeron 10 días después de la electroporación, se trataron con tripsina y se cultivaron en placas de 96 pocillos gelatinizados durante al menos tres días antes de la extracción y purificación de ADN. Los clones de células ES dirigidos correctamente se identificaron mediante el ensayo de pérdida de alelos (Frendewey *et al.* (2010), *Methods in enzymology* 476:295-307; Valenzuela *et al.* (2003a), *Nat. Biotechnol.* 21: 652-659).

Ejemplo 3: Preparación de ratones ARNIinc

Se utilizó el método VelociMouse® (Dechiara, T.M., (2009), *Methods Mol Biol* 530:311-324; Lumeng *et al.* (2007), *Nat. Biotechnol.* 25:91-99), en el que las células ES se inyectaron en embriones Swiss Webster en la etapa de 8 células no compactadas para producir una generación F0 totalmente derivada de células ES que llevan las mutaciones de inactivación ARNIinc. Se usaron VelociMice® machos directamente para el perfil de expresión de *lacZ* o se aparearon con hembras C57BL/6NTac para producir embriones o adultos para el análisis de *lacZ* o para producir criadores de F1 y se realizaron estudios fenotípicos sobre los ratones de N2F1. Los emparejamientos programados se llevaron a cabo asignando la mañana de identificación de los tapones vaginales como día 0,5 (E0,5).

Ejemplo 4: Perfil de expresión de *LacZ*

Para la tinción de montaje completo, se recogieron embriones de E9,5 y E14,5, se lavaron en PBS y se incubó durante 15 a 60 minutos en solución de glutaraldehído al 0,2 %. Los sacos de yema de embrión se sacaron para el genotipado. Después de la fijación, se lavaron los embriones en tampón de lavado y se incubaron en solución de tinción X-gal (1 mg/ml) a 37 °C de 1 a 24 horas. Después de la tinción, los tejidos se aclararon en tampón de lavado, se fijaron posteriormente en paraformaldehído al 4 % y se incubaron en etanol al 70 % durante al menos 24 horas. Se fotografiaron los embriones E9,5-e11,5 inmediatamente mientras los embriones e12,5 y mayores se aclararon en una serie de soluciones de contenido en glicerol creciente y de KOH al 1 % decreciente en ddH₂O. Las fotografías se tomaron con un microscopio estereoscópico Nikon SMZ800. Se diseccionaron los pulmones de los embriones *Fendrr* e13,5 para fotografiarlos después del aclarado.

Para los estudios con ratones adultos, se anestesiaron en profundidad VelociMice® derivados de células ES completamente de generación F0 de 6 a 8 semanas de edad y se fijaron mediante perfusión cardiaca utilizando una solución de glutaraldehído al 0,2 % /paraformaldehído al 4 %. Se diseccionaron tejidos de cerebro, caja torácica, corazón, pulmón, hígado, bazo, estómago, riñón, intestino, urogenital, músculo y extremidades traseras, se aclararon en PBS y se fijaron posteriormente durante 30 minutos en una solución de glutaraldehído al 0,2 % /paraformaldehído al 4 %. A continuación, los tejidos se lavaron y se incubaron en solución de tinción X-gal 1 mg/ml) durante 1 a 24 horas a 37 °C. Después de la tinción, los tejidos se aclararon en tampón de lavado, se fijaron posteriormente en paraformaldehído al 4 %, se aclararon en una serie de glicerol al 50 %, 70 % y 100 % y se fotografiaron como se hizo para los embriones.

Ejemplo 5: Cuidado animal y procedimientos experimentales

Los estudios fenotípicos de ratones N2F1 comenzaron a las 6-8 semanas de edad. Para los emparejamientos programados, se asignó la mañana de identificación de los tapones vaginales como día embrionario 0,5 (E0,5). Los compañeros de camada KO de ARNlinc y de tipo silvestre se observaron desde el nacimiento para diversas metas del desarrollo (enanismo, respiración, anomalías faciales y de patas, color de piel, postura, enderezamiento y apertura de ojos) hasta aproximadamente 6-8 semanas de edad, cuando se les colocó en 12 h de luz al día a 69-74 F y el 40-60 % de humedad para el estudio. Todos los experimentos comenzaron a las 6-8 semanas de edad y todos los procedimientos animales se realizaron en cumplimiento de los protocolos aprobados por el Comité institucional de cuidado y uso de animales (IACUC; del inglés, Institutional Animal Care and Use Committee) de Regeneron Pharmaceuticals.

Ejemplo 6: Análisis de μ CT

Las imágenes esqueléticas 3D se obtuvieron utilizando el sistema de imágenes preclínicas *in vivo* de microCT de Quantum FX (Perkin Elmer). Los ratones se anestesiaron usando la inhalación de oxígeno/isoflurano con una tasa de flujo de isoflurano de 2,5 L/min y una tasa de flujo de oxígeno de 1,5 L/min. Durante la exploración, la anestesia se mantuvo en una tasa de flujo de oxígeno de 0,25 L/min a través de un cono nasal. Las exploraciones se realizaron a 90kV y 160 μ A con un campo visual de 30 mm de visión para las patas posteriores y un campo visual de 60 mm para las vértebras. Para la densidad mineral ósea, los volúmenes de hueso total, magro y grasa, se realizaron dos exploraciones consecutivas con un campo visual de 60 mm para el cuerpo completo, excluyendo la cabeza. El fémur derecho se aisló manualmente para las mediciones de densidad mineral ósea. Los volúmenes de fémur derecho, magro total y grasa total se midieron todos utilizando el software Analyze 11.0 (Mayo Clinic) y se convirtieron en masa basándose en densidades establecidas. Tras la exploración, los ratones se devolvieron a su jaula y se monitorizaron para su recuperación en cumplimiento de los protocolos del IACUC de Regeneron.

Ejemplo 7: Prueba de suspensión de la cola

Cuando se suspenden por la cola, los ratones se preparan para un aterrizaje seguro expandiendo sus patas traseras, a menudo referido como "ensanchamiento de patas traseras". Se suspendió a los ratones por la cola durante 10s y se observó cualquier fenotipo de juntado.

Ejemplo 8: Prueba de resistencia de prensión

Los ratones se evaluaron a las 5, 7 y 10 semanas de edad en cuanto a signos de déficit muscular por su capacidad para sostenerse invertidos desde una rejilla de alambre (grosor del alambre de 2 mm aproximadamente). Los ratones se colocaron individualmente sobre una rejilla metálica que fue agitada suavemente para incitarles a mantenerse según la rejilla se ponía boca abajo. Se grabó el tiempo para que el ratón se soltara (hasta un máximo de 60 segundos). Se les concedieron a los ratones tres intentos de mantenerse el mayor tiempo posible y el tiempo máximo se registró para la comparación estadística.

Ejemplo 9: Histología muscular y necropsia tisular

Los ratones se sacrificaron por inhalación de CO₂ seguida de dislocación cervical. Se diseccionaron y pesaron los músculos tibial anterior (TA) y gastrocnemio (GA). Todos los músculos y órganos recogidos se congelaron y se mantuvieron a -80C para examinación futura. Para la histología, los músculos se congelaron en OCT, se crioseccionaron transversalmente en espesores de 12 μ m para revelar la cabeza lateral y medial, sóleo y plantar. Las secciones adyacentes se tificaron con H&E, laminina y tinción lenta CMH (en inglés, MHC). Las tinciones se fotografiaron digitalmente utilizando un Aperio Scanscope. El tamaño y recuento de fibras se determinó utilizando el software Spectra. Todos los datos se expresan como medias +/- error típico de la media (ETM) (representado como barras de error). El análisis de varianza (ANOVA) se realizó utilizando el programa STATVIEW y/o PRISM. La significación estadística se estableció como un p valor de menos de 0,05. Para la histología de piel, se rasuraron áreas de piel dorsales y ventrales, se diseccionaron y fijaron en paraformaldehído al 4 % (PFA) durante al menos 24 horas y se transfirieron a etanol al 70 %. La inclusión en parafina, el corte y la tinción con hematoxilina y eosina sobre secciones de piel se realizaron por los laboratorios Histoserv Labs, Inc., Germantown, Maryland.

Ejemplo 10: Análisis de curva de supervivencia Kaplan-Meier

Se observó a los animales durante un periodo de 52 semanas y se monitorizaron para detectar signos de morbilidad según los protocolos del IACUC de Regeneron. Ningún ratón en este estudio necesitó ser sacrificado antes del punto temporal de 52 semanas basándose en pautas de morbilidad. Las curvas de supervivencia y la prueba de rangos logarítmica se determinaron utilizando el software Graphpad PRISM 6.

Ejemplo 11: Fenotipos diversos y patrones de transcripción específicos en veinte líneas de ratones con ARNlinc ablacionado

En una investigación de 20 líneas de ratones nuligénicos diseñadas para examinar las funciones biológicas de ARN no codificantes intergénicos grandes (ARNlinc), los investigadores han encontrado diversos fenotipos, variando desde letalidad perinatal a defectos asociados con envejecimiento prematuro y anomalías morfológicas y funcionales en los pulmones, esqueleto y músculo. Cada alelo mutante llevó un indicador de *lacZ* cuyo perfil de expresión destacó un amplio espectro de patrones espaciotemporales y de transcripción específicos de tejidos en embriones y adultos que resultaron informativos para los análisis fenotípicos y que servirán para investigaciones futuras de estos genes. El presente estudio muestra que los ARNlinc son una nueva clase de moléculas codificadas que, como las proteínas, cumplen roles funcionales esenciales e importantes en el desarrollo embrionario, fisiología y homeostasis de una amplia matriz de tejidos y órganos en mamíferos.

Recientemente ha quedado claro que una comprensión profunda de la relación entre genotipo y fenotipo en los mamíferos requiere que se amplíen nuestras investigaciones más allá de los genes que codifican proteínas para incluir la porción no codificante del genoma (Mattick JS (2009) PLoS genetics 5: e1000459). Los estudios a gran escala de expresión del genoma completo en células de mamífero han revelado que aproximadamente tres cuartas partes del genoma es capaz de expresarse como ARN (Kapranov, *et al.* (2007) Science 316: 1484-1488; Carninci P, *et al.* (2005) Science 309: 1559-1563; Djebali S, *et al.* (2012) Nature 489: 101-108) y la mayoría de los transcritos no codifican para proteínas. Entre los transcritos no codificantes existe una clase diversa conocida como los ARNlinc no codificantes (ARNlinc). Representando aproximadamente 15.000 transcritos de los cerca de 10.000 loci genómicos en células humanas (Derrien, T., *et al.* (2012), Genome Res 22: 1775-1789.), los ARNlinc y una subclase conocida como ARN no codificante intergénico grande (ARNlinc) (Guttman, M., *et al.* (2009), Nature 458: 223-227; Khalil AM, *et al.* (2009) Proceedings of the National Academy of Sciences of the United States of America 106: 11667-11672) se asemejan a ARNm codificante de proteínas en estructura, síntesis y el carácter cromatínico de sus genes. Si esta similitud estructural se extiende hacia una diversidad funcional que empareje proteínas o no, permanece como una cuestión abierta.

Desde la creación de la primera cepa nuligénica hace casi veinticinco años, el ratón ha llegado a ser el sistema más importante para el estudio de la función génica en mamíferos (Capecchi, M.R. (2001) Nat Med 7: 1086-1090; Evans MJ (2001) Nat Med 7: 1081-1083; Smithies O (2001) Nat Med 7: 1083-1086). Con pocas excepciones, la aplicación de tecnología de ratones nuligénicos en estudios génicos individuales así como en proyectos internacionales a gran escala (www.knockoutmouse.org) se han centrado en genes codificantes de proteínas, pero los esfuerzos recientes para crear recursos de ratones nuligénicos globales para microARN (Prosser, *et al.* (2011) Nature biotechnology 29: 840-845) (mcmanuslab.ucsf.edu/microna_knockout) demuestran el valor de aplicar la tecnología de ARN no codificantes. Ha habido unos pocos estudios funcionales de ARNlinc individuales mediante alteración génica en ratones, pero aproximadamente la mitad se han centrado en los ARNlinc bien estudiados implicados en un fenómeno biológico relacionado sencillo: la inactivación del cromosoma X chromosome (Marahrens Y, *et al.* (1997) Genes & Development 11: 156-166; Sado T, *et al.* (2001) Development 128: 1275-1286) e impronta cromosómica somática (Leighton PA, *et al.* (1995) Nature 375: 34-39; Mohammad F, *et al.* (2010) Development 137: 2493-2499; Sleutels F, *et al.* (2002) Nature 415: 810-813; Takahashi N, *et al.* (2009) Human Molecular Genetics 18: 1879-1888).

Recientemente, la alteración del ARNlinc *Fendrr* de ratón dio como resultado una letalidad embrionaria asociada con defectos en el desarrollo del corazón y de la pared corporal (Grote P, *et al.* (2013) Developmental Cell 24: 206-214). Sin embargo, las mutaciones de delección o inserción en los genes Gt(ROSA)26Sor (Zambrowicz BP, *et al.* (1997) Proceedings of the National Academy of Sciences of the United States of America 94: 3789-3794) o Malat1 (Zhang B, *et al.* (2012) Cell Reports 2: 111-123) que codifican ARNlinc produjeron fenotipos no discernibles. La comprensión emergente de la estructura, expresión y función de los genes de ARNlinc presenta una nueva oportunidad para emplear genéticas moleculares de ratón para revelar las funciones biológicas asociadas con esta nueva clase de genes.

La aplicación de tecnología de ratones nuligénicos a ARNlinc, sin embargo, presenta algunos desafíos técnicos. La mayoría de las proteínas tienen elementos o dominios que son conocidos por ser o al menos se prevén importantes para funcionar. La delección de secuencias codificantes para estas partes esenciales a menudo es suficiente para crear un alelo nulo. Asimismo, se pueden diseñar alelos condicionales que aislen el exón o exones críticos para una delección posterior mediante la acción de una recombinasa específica de tejidos. Debido a que no se han establecido todavía relaciones estructura-función para todos menos para algunos ARNlinc y a que no hay marco de lectura abierto como una guía, las estrategias de inactivación disponibles para genes codificantes de proteínas pueden no ser aplicables a los loci genómicos que codifican ARNlinc. Aunque la anotación de los genes de ARNlinc ha mejorado (Derrien T, *et al.* (2012) Genome Research 22: 1775-1789), los límites precisos de algunos genes todavía permanecen ambiguos, lo

que puede complicar el diseño de alelos de inactivación. Una herramienta potente para aplicar a ratones nuligénicos para genes codificantes de proteínas es el reemplazo del gen diana con un indicador, tal como la secuencia codificante para β -galactosidasa o una proteína fluorescente, cuya expresión está controlada por el promotor del gen diana, informando de este modo del patrón espacial y temporal de su expresión en el ratón. El reemplazo de genes indicadores se ha aplicado exitosamente a ARN no codificante, tal como el locus bien estudiado Gt(ROSA)26Sor (Zambrowicz, B.P., *et al.* (1997), Proc Natl Acad Sci USA 94: 3789-3794), que codifica un ARNlinc y el gen para el gen para el ARN no codificante pequeño miR-155 (Thai, T.H., *et al.* (2007) Science 316: 604-608), pero se necesitaría desarrollar reglas para crear tales alelos para ARNlinc. A pesar de estas cualificaciones, con miles de ARNlinc identificados, ha llegado el momento de aplicar la potencia de la tecnología de ratones nuligénicos para esta nueva clase de genes. Con este objetivo en mente, los inventores describen en el presente documento un enfoque genético unificado para dilucidar las funciones de veinte ARNlinc mediante la creación de líneas de ratones nuligénicos, cada una portando un alelo de delección de ablación génica con un reemplazo de indicador de β -galactosidasa.

Generación de 20 líneas de ratones de ARNlinc suprimidos con reemplazo de genes indicadores

La Tabla 1 enumera los 20 genes de ARNlinc en 10 cromosomas diferentes que fueron objetivo en este estudio y los 26 alelos de delección de inactivación creados. Los presentes inventores eligieron mutar miembros de la clase de ARN no codificante intergénico grande, por definición, los genes de ARNlinc se aíslan de genes codificantes de proteínas vecinas y sus transcritos no solapan (Guttman M, *et al.* (2009) Nature 458: 223-227). Esta característica nos permitió diseñar alelos de delección que tendrían menos posibilidades de interferir con la expresión de genes cercanos. Los presentes inventores eligieron genes de ARNlinc dirigidos para reflejar diversos patrones de expresión (Khalil, A.M., *et al.* (2009), Proc Natl Acad Sci USA 106: 11667-11672; Cabili MN, *et al.* (2011) Genes Dev 25: 1915-1927), con énfasis en la expresión neuronal y para su implicación potencial en el desarrollo y la regulación de la expresión génica.

La estrategia de diseño de los presentes investigadores para las mutaciones de inactivación de ARNlinc estuvo guiada por dos objetivos. En primer lugar, los presentes inventores tenían como objetivo crear alelos que informaran con precisión de la actividad de transcripción de los genes ARNlinc. Aunque existía evidencia a partir de los estudios basados en células y de disección de los tejidos seleccionados para la expresión de ARNlinc específicos de tejidos (Cabili MN, *et al.* (2011) Genes & Development 25: 1915-1927), los presentes inventores quisieron complementar esta base de conocimiento produciendo los patrones de expresión de mayor definición ofrecidos por el perfil de expresión de *lacZ*, que puede resolver la expresión de tejidos y órganos tanto espacial como temporalmente y de este modo, revelar subdominios y en algunos casos, la especificidad de tipo de célula no resuelta mediante experimentos de disección de tejidos. Segundo, los presentes inventores se esforzaron por crear delecciones ablativas de genes que suprimieron la síntesis y la función del ARNlinc de modo que cualesquiera fenotipos asociados con las mutaciones serían informativos sobre las funciones críticas de los ARN dirigidos.

Las delecciones de inactivación variaron en tamaño de aproximadamente 400 pb a 50 kb, con la mitad eliminando todos los exones anotados. Para la mayoría de los alelos restantes, la delección comenzó en el segundo exón. La aplicación de los métodos VelociGene® (Valenzuela, D.M., *et al.* (2003), Nat Biotechnol 21: 652-659) para la construcción y uso de vectores de direccionamiento grandes basados en cromosomas artificiales bacterianos (LTVEC) fue crucial permitir a los presentes inventores hacer las delecciones de ablación de genes grandes necesarias para asegurar un alelo nulo para esta nueva clase de ARN funcional grande.

Poco se sabe acerca de la relación entre estructura y función de los genes ARNlinc que pueden guiar el diseño de alelos. La experiencia con la alteración de los genes Gt(ROSA)26Sor (Zambrowicz *et al.*, (1997) Proceedings of the National Academy of Sciences of the United States of America 94: 3789-3794) y BIC (miR-155) (Thai *et al.* (2007) Science 316: 604-608), estableció que la delección e inserción después del primer exón puede producir una expresión fiable y específica de tejidos de la β -galactosidasa u otros indicadores. Esta estrategia puede, sin embargo, fallar en conseguir una mutación nula completa si el transcrito de fusión a partir del alelo modificado retiene una parte funcional del ARNlinc de la porción 5' codificada en el primer exón (Tsai MC, *et al.* (2010) Science 329: 689-693). Los diseños de alelos de inactivación indicados en la Tabla 1 fueron, por tanto, un compromiso entre el deseo de una mutación completamente ablativa que tuviera la mayor probabilidad de suprimir la función de ARNlinc y el objetivo de crear un alelo que produjera un perfil de expresión de genes preciso e informativo a partir del indicador de β -galactosidasa. Por ejemplo, para el gen *HOTAIR* los presentes inventores hicieron dos alelos, uno que suprimió casi la secuencia codificante de ARN entera y un segundo en el que la delección se inició en el segundo exón. Ambos alelos produjeron fenotipos idénticos (descritos más adelante), pero solo el segundo funcionó como un indicador de expresión génica.

Para ARNlinc que residen muy cerca de un gen codificante de proteínas y pueden compartir un promotor divergente, los presentes inventores establecieron el punto de inicio de delección se estableció en el segundo exón para evitar la posibilidad de alterar la transcripción del gen vecino. La Figura 1 muestra tal ejemplo para el gen *Fendrr*. El diagrama muestra los elementos de diseño comunes a todos los alelos: una delección dirigida de toda o la mayoría de la secuencia codificante para el ARNlinc y reemplazo con un casete que contiene una secuencia del gen *lacZ* de *E. coli* que codifica β -galactosidasa y un casete (*neof*) que expresa neomicina fosfotransferasa para la selección de colonias de células ES resistentes a G418. Los sitios de reconocimiento de recombinasa de *LoxP* que permiten la escisión mediada por Cre antes del análisis fenotípico flanquean el casete de selección de fármacos. Como no hay lectura abierta funcional con la que fusionar la secuencia *lacZ*, cada alelo lleva un codón de iniciación y una secuencia consenso Kozak (Kozak,

M. (1987), *Nucleic Acids Res* 15, 8125-8148) para la traducción eficaz del indicador de β -galactosidasa.

Patrones de expresión génica de ARNlinc específicos y diversos revelados por el perfilado del indicador *LacZ*

5 Para investigar los patrones de expresión de los 20 genes de ARNlinc dirigidos, los presentes investigadores aplicaron tinción X-gal para la actividad de β -galactosidasa para embriones en la etapa media de gestación y tejidos y órganos montados completos de adultos. Los genes de ARNlinc dirigidos mostraron diversos patrones de expresión génica de indicadores tanto en embriones como en adultos, representando la mayoría de los principales sistemas de órganos y tipos de tejido (Tabla 2 en la Figura 11). Por ejemplo, en los tejidos de adultos, la expresión de *Pantr2*, *Kantr* y *Peril* se limitó al cerebro; *Mannr* y *Fendrr* se expresaron en pulmones; *Eldr* se expresó en el sistema urogenital; y *Halr1* se expresó en la caja torácica. Un gen de ARNlinc, *Pint*, mostró una expresión ubicua en todos los tejidos. Los presentes inventores no detectaron expresión de los genes *Hotair*, *Ptgs2os2* y *Haglr* en ninguno de los tejidos de adultos examinados.

15 La expresión embrionaria parece ser una característica común de los ARNlinc. El examen de la expresión del indicador de β -galactosidasa en embriones heterocigóticos el día o alrededor del día embrionario 12,5 (E12,5) reveló diversos patrones específicos para todos los 20 genes de ARNlinc dirigidos (Tabla 2 en la Fig. 11, Fig. 2A). Los perfiles de expresión variaron desde ubicuos (*Tug1*) a altamente específicos, tales como epidérmico para *Eldr*, placoda de bigotes para *Trp53cor1* (Fig. 9) o los brotes mamarios para *lincenc1* (Fig. 9). Los patrones espaciotemporales observados en los diferentes alcances de la expresión de brote de extremidades y cola para *Hottip* y *Hoxallos* son muy similares a aquellos informados para los genes codificantes de proteínas adyacentes en el grupo HoxA (Hostikka SL, Capecchi MR (1998) *Mechanisms of Development* 70: 133-145; Lu P, *et al.* (2008) *Development* 135: 1395-1405). La expresión de *Hotair* en el brote posterior de cola y tubérculo genital que los presentes investigadores observaron para el indicador de β -galactosidasa fue idéntica a la determinada mediante hibridación *in situ* (Schorderet, P, Duboule D (2011) *PLoS genetics* 7: e1002071). El análisis de la tinción de X-gal en diferentes puntos durante el desarrollo embrionario mostró que para algunos de los genes ARNlinc, la expresión comenzó de forma temprana en un sitio restringido y a continuación, se extendió más allá de este locus inicial en etapas posteriores (Fig. 2B), de nuevo, recuerda a la expresión de proteínas de Hox (Nagy, A. (2003) *Manipulating the mouse embryo: a laboratory manual*. 3ª ed. Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press. pp. x, 764 p.). Por ejemplo, la expresión de los genes *Hottip* y *Hoxallos* comenzó en el extremo posterior del embrión de E9,5 y a continuación, se extendió en los brotes de extremidades en momentos posteriores. De forma análoga, la expresión inicial de *Celrr* en un sitio cerca del extremo anterior de los embriones en E9,5 se mantuvo y se expandió a toda la longitud del tubo neural en los siguientes dos días.

35 En consonancia con la expresión cerebral frecuente observada entre los ARNlinc específicos de tejidos humanos (Derrien *et al.* (2012) *Genome Research* 22: 1775-1789), los presentes investigadores encontraron que la mitad de los 20 genes de ARNlinc de ratón dirigidos son activos transcripcionalmente en el cerebro adulto. Como con la expresión de ARNlinc embrionaria, los patrones cerebrales (Fig. 3) resultaron únicos y variaron desde ubicuos (*lincppara* y *Pint*) a estructuras cerebrales específicas altamente restringidas (*Peril*, *Crnde* y *Kantr*).

40 La expresión incrementada única de *Pint* con la edad se correlaciona con un fenotipo de tipo envejecimiento

De los 20 genes de ARNlinc dirigidos, solo *Pint* mostró un patrón de expresión corporal completo, restringido principalmente a la vida posnatal (Tabla 2 en la Figura 11). Exclusivo de *Pint*, los presentes inventores observaron un incremento de su expresión con la edad (Fig. 4). En neonatos de 3 días, la actividad transcripcional de *Pint* fue baja (cerebro) o indetectable (músculo de la caja torácica), pero a continuación apareció gradualmente en ratones de 3 semanas de edad y se volvió fuerte y ubicua a las 8 semanas de edad. Aunque la fuerza y el momento de expresión de *Pint* varió entre los distintos órganos y tejidos, la tendencia general fue un incremento constante en la expresión después del nacimiento a una meseta en la edad adulta. Según el conocimiento de los inventores, este patrón de expresión dinámico relacionado con la edad es nuevo. Los presentes inventores no observaron un perfil similar en su experiencia con experimentos de perfilado de *lacZ* para cientos de genes codificantes de proteínas (Valenzuela DM, *et al.* (2003) *Nature Biotechnology* 21: 652-659).

El sorprendente incremento relacionado con la edad en la expresión de *Pint* corporal completa revelado por el perfilado de *lacZ* (Fig. 4) sugirió un papel homeostático global para *Pint* en el mantenimiento de la salud normal según envejecen los ratones. Para probar esta hipótesis, los presentes investigadores criaron la línea de ratones nuligénicos para *Pint* hacia la homocigosis y realizaron un estudio longitudinal que comparó ratones homocigóticos (*Pint*^{-/-}) con los de tipo silvestre (WT) y los controles heterocigóticos (*Pint*^{+/-}) de compañeros de camada. Los ratones *Pint*^{-/-} se mostraron sanos y normales al nacer; sin embargo, a la edad de 3 meses comenzaron a mostrar signos de un inicio temprano de un fenotipo similar a envejecimiento. Las mediciones de peso corporal revelaron que tanto machos como hembras de ratones *Pint*^{-/-} mostraron una tasa de crecimiento más lenta en comparación con sus compañeros de camada WT, pero fue más pronunciado en los machos (Fig. 5A). Con un año de vida, los ratones macho *Pint*^{-/-} fueron un 30 % más livianos y los ratones *Pint*^{+/-} fueron un 15 % más livianos que sus compañeros de camada WT, mientras que las hembras *Pint*^{-/-} fueron un 27 % más ligeras (datos no mostrados). Los análisis Kaplan-Meier que comparan los ratones macho homocigóticos con los heterocigóticos y WT (Fig. 5B) demostraron que la pérdida de *Pint* se asocia con un resultado de escasa supervivencia. Los presentes inventores no encontraron signos de tumores o lesiones en los

ratones mutantes a medida que envejecían, pero algunos ratones *Pint*^{-/-} desarrollaron hernia, incluyendo una protrusión del proceso xifoides del pecho asociado con adelgazamiento de la pared abdominal (datos no mostrados). Hubo una postura de juntado de las patas traseras anormal dependiente de la edad cuando los ratones se suspendieron de la cola (datos no mostrados). La gravedad de este fenotipo varió, pero su frecuencia se incrementó progresivamente con la edad, sugiriendo un declive de la fuerza muscular (véase la Fig. 8 para otro ejemplo en la línea nuligénica para *Hottip*). Los presentes inventores también observaron pérdida de pelaje en ratones tanto machos como hembras (datos no mostrados). Los análisis histológicos de las secciones de piel recogidas de los cuerpos ventrales y dorsales de ratones *Pint*^{-/-} revelaron fibrosis y una diferencia considerable en el desarrollo del folículo piloso junto con una reducción dramática en espesor de la capa subcutánea de grasa (Fig. 5C).

Los análisis no invasivos de cuerpo completo mediante microtomografía de rayos X (microCT) de ratones individuales a medida que envejecían indicaron un contenido de grasa significativamente más bajo en ratones *Pint*^{-/-} machos (Fig. 5D) y hembras (Fig. 5E) en comparación con sus compañeros de camada WT. La pérdida total de grasa corporal probablemente fue la principal contribución al declive en peso corporal a medida que envejecían (Fig. 5A). Los ratones *Pint*^{-/-} también tuvieron una densidad mineral ósea del fémur significativamente más baja que los WT (Fig. 5F y 5G). Los ratones macho tuvieron una masa magra significativamente disminuida en las 52 semanas de edad. Tanto machos como hembras mostraron una masa muscular significativamente disminuida para el complejo gastrocnemio (GA) y el tibial anterior (TA) que comienza en las 26 semanas de edad (no mostrado). Las imágenes esqueléticas revelaron la aparición de una cifolordosis grave en ratones *Pint*^{-/-} tanto machos como hembras en comparación con los WT (Fig. 5H). Aproximadamente un 70 % de los ratones *Pint*^{-/-} de 12 semanas de edad mostraron cifolordosis y un 100 % a las 52 semanas de edad (Fig. 5I). En cambio, solo un 10 a un 20 % de los ratones WT de 26 semanas de edad mostraron una ligera cifolordosis y esta frecuencia no se incrementó con la edad. Los ratones *Pint*^{-/-} no desarrollaron una cifolordosis significativa hasta las 26 semanas de edad, indicando una haploinsuficiencia dependiente de la edad inusual para *Pint*. El espectro de patologías asociadas con la edad en los ratones nuligénicos para *Pint* sugieren que *Pint* puede ser importante para el mantenimiento de la salud y la evitación del envejecimiento prematuro durante el ciclo de vida normal del ratón.

La pérdida de *Fendrr* provoca letalidad perinatal como resultado de una dificultad respiratoria

De las 20 líneas de ratones nuligénicos para ARNlinc, los ratones *Perit*^{-/-} y *Fendrr*^{-/-} mostraron letalidad perinatal. El alelo inactivado de *Fendrr* utilizado por los presentes investigadores tiene una delección de 26 kb desde el exón 2 al último exón anotado (Fig. 1). La tinción de X-gal de embriones homocigóticos de E12,5 mostró una expresión de *lacZ* en el proceso frontonasal, tracto respiratorio superior, pulmones y en la región aorta gónada mesonefros (AGM) posterior (Fig. 6A) que fue idéntica en embriones tanto heterocigóticos (no mostrado) como homocigóticos, indicando una organogénesis sumamente normal. Una mirada aislada a los pulmones en desarrollo en la E13,5 reveló defectos en los embriones nuligénicos: los pulmones fueron pequeños y los lóbulos parecieron globulares y desorganizados (Fig. 6B). Los ratones homocigóticos para la delección del gen *Fendrr* sobrevivieron al nacimiento pero sucumbieron poco después por problemas respiratorios evidentes. El fenotipo letal perinatal mutante *Fendrr* fue idéntico en ratones sobre 2 antecedentes genéticos diferentes: un híbrido C57BL6/129 informado aquí y en ratones adicionalmente retrocruzados con C57BL6 en un programa de cría separado (Sauvageau M, *et al.* (2013) *Elife* 2: e01749).

La pérdida de *Hotair* y *Hottip* provoca defectos morfológicos y funcionales en esqueleto y músculo

La tinción de X-gal embrionaria para los genes *Hotair* y *Hottip* mostró una expresión restringida en los brotes de extremidades posteriores y distales (Fig. 2A). En consonancia con estos patrones de expresión restringidos por el desarrollo, las delecciones de los genes *Hotair* y *Hottip* provocaron malformaciones morfológicas en la cola y extremidades traseras de ratones adultos. En ratones *Hotair*^{-/-} los presentes investigadores observaron una transformación homeótica evidente de la 4^a vértebra caudal, que se volvió anatómicamente similar a la 3^a vértebra caudal (Fig. 7). Los ratones *Hottip*^{-/-} mostraron una postura de juntado de las patas traseras anormal cuando se les suspendió de la cola en comparación con sus compañeros de camada de tipo silvestre (Fig. 8A). Esta anomalía de comportamiento se acompañó de una pérdida en la resistencia de prensión como se midió mediante una prueba en la que se desafiaba a los ratones a permanecer suspendidos en una jaula de alambre invertida. El tipo silvestre y los mutantes *Hottip*^{-/-} se mantuvieron durante aproximadamente un minuto, mientras que sus compañeros de camada homocigóticos liberaron su prensión en 10-20 segundos (Fig. 8B). Esta reducción evidente de la fuerza de prensión se asoció con una pérdida de masa muscular del gastrocnemio, pero no del tibial anterior o del cuádriceps (Fig. 8C). Los presentes inventores observaron una reducción aproximada del 40 % en el número de fibras musculares en el gastrocnemio pero no una reducción en el tamaño de fibra promedio (Fig. 8D y E). Además de los defectos musculares en los ratones nuligénicos para *Hottip*, los presentes investigadores también encontraron una malformación esquelética de extremidades traseras: un acortamiento en la longitud del hueso calcáneo (Fig. 10).

En los últimos años ha habido una explosión en la comprensión de los presentes inventores del componente no codificante de proteínas del genoma, especialmente en mamíferos. Además de las clases de ARN funcionales no codificantes reconocidas desde hace mucho tiempo tales como ARN ribosómico, de transferencia, pequeño nuclear, pequeño nucleolar, pequeño citoplasmático, los componentes de ARN de la ARNasa P, ARNasa MRP y enzimas telomerasas y los descubiertos más recientemente microARN y los ARNpi asociados a PIWI, se pueden incluir ahora al menos 15.000 miembros de la clase ARN no codificante largo (Kapranov, *et al.* (2007) *Science* 316: 1484-1488;

Carninci P, *et al.* (2005) *Science* 309: 1559-1563; Djebali S, *et al.* (2012) *Nature* 489: 101-108; Derrien T, *et al.* (2012) *Genome Research* 22: 1775-1789; Guttman M, *et al.* (2009) *Nature* 458: 223-227). Según se va entendiendo la presencia genómica y la expresión de genes ARNlinc, el siguiente objetivo es descubrir sus funciones biológicas. Como una primera etapa para abordar este reto, los presentes investigadores aplicaron la tecnología de direccionamiento genético de ratones, la herramienta más potente para la determinación de la función génica en mamíferos, para crear un recurso de líneas de ratones nuligénicos para 20 genes de ARNlinc (Sauvageau M, *et al.* (2013) *Elife* 2: e01749).

Las relaciones de estructura-función para los ARNlinc se conocen poco. Por este motivo, fue crucial en este estudio inicial crear alelos de inactivación con deleciones que suprimieron la mayoría si no todo el potencial codificante de ARNlinc, para asegurar tener la mayor probabilidad de crear una mutación de pérdida de función. La anotación ambigua e incompleta de muchos loci de ARNlinc, con múltiples transcritos informados, tal vez generados mediante corte y empalme alternativo o sitios de inicio de la transcripción, se añade a la dificultad de diseñar alelos de inactivación. La nueva comprensión de las características moleculares importantes para la función de ARNlinc debe informar del diseño de la siguiente generación de alelos de ARNlinc con modificaciones dirigidas de forma más precisa de secuencias críticas para funcionar y permitir también estrategias condicionadas avanzadas y flexibles.

Un objetivo clave de la investigación de inactivación de ARNlinc de los presentes investigadores el presente documento fue crear alelos que además de abolir una función también indicaron el patrón espaciotemporal de expresión del gen. A pesar de no tener un marco de lectura abierto como una guía, los presentes inventores diseñaron alelos satisfactoriamente que informaron de expresión génica de todos los 20 genes dirigidos. Uno de los alelos que no produjo expresión de *lacZ* en la etapa adulta fue *Ptgs2os2* (véanse la Fig. 2A y la Fig. 9 para la expresión embrionaria), que se conoce por ser uno de los ARNlinc más fuertemente inducidos por señales inflamatorias (Guttman M, *et al.* (2009) *Nature* 458: 223-227; Carpenter S, *et al.* (2013) *Science* 341(6147): 789-92). La línea de inactivación *Ptgs2os2* debe demostrar ser un recurso valioso para estudios de cómo una expresión de ARNlinc responde a una infección u otros ataques inflamatorios y qué papel biológico juega en el proceso.

Uno de los criterios que los presentes inventores aplicaron en su selección de qué genes de ARNlinc dirigir para esta investigación fue una expectativa de expresión en tejidos neuronales. Diez de los genes dirigidos mostraron una expresión del indicador *lacZ* en el cerebro de adultos y cada uno mostró un patrón único (Fig. 3), variando desde fuerte expresión en todo el cerebro (*Pint*) a ligera expresión de la materia gris en la mayoría de estructuras (*Tug1*) a expresión altamente restringida exclusiva de los colículos (*Crnde*) o de la línea media del hipotálamo (*Peril*). La variedad y especificidad de los patrones de expresión génica en el cerebro fue también evidente en otros tejidos y fue similar a aquellos que los presentes investigadores habían observado con los alelos indicadores para los genes codificantes de proteínas. Los patrones de perfil de expresión del gen *lacZ* de ARNlinc que los presentes investigadores encontraron fueron coherentes con la expresión específica de tejido encontrada mediante los experimentos de cuantificación de ARN en tejidos de ratón de tipo silvestre (Sauvageau M, *et al.* (2013) *Elife* 2: e01749). Antes de este estudio, sin embargo, la especificidad exquisita tisular y del tipo de célula de la expresión génica del ARNlinc no se apreció, debido a que los métodos de cuantificación previos no pudieron suministrar la alta definición y resolución del tipo de células del perfilado del indicador *lacZ* (Fig. 2A).

La expresión embrionaria fue una característica compartida por todos los genes ARNlinc que los presentes inventores examinaron. El perfilado de *lacZ* suministró una visión de alta definición de embriones completos que reveló el amplio intervalo de patrones específicos únicos para cada ARNlinc. Entre los ejemplos se incluyen la expresión específica exquisita observada en la placoda de bigotes para *Trp53cor1* y el brote mamario para *lincenc1*, la expresión epidérmica de *Eldr*, la expresión de brote de extremidades de *Hottip* y *Hoxallos* y la expresión ubicua de *Tug1* (Fig. 2A y Fig. 2B y Fig. 9). Estos patrones diversos podrían señalar un papel común para los ARNlinc en la regulación de eventos clave en el desarrollo.

Otro valor del perfilado de *lacZ* es que puede guiar y enfocar el diseño de estudios fenotípicos. Por ejemplo, los patrones de expresión posteriores altamente restringidos para los embriones *Hotair* y *Hottip* sugirieron que los presentes investigadores podrían encontrar fenotipos nuligénicos en partes del cuerpo posteriores. En coherencia con esta expectativa, los presentes inventores observaron una transformación homeótica evidente de la 4ª vértebra caudal en ratones *Hotair*^{-/-} (Fig. 6) y encontraron anomalías de las patas traseras que incluyeron debilidad muscular y malformaciones esqueléticas en ratones *Hottip*^{-/-} (Fig. 8 y Fig. 10). El fenotipo homeótico *Hotair* se ha observado también en ratones con un alelo de inactivación de *Hotair* diferente (H. Chang, comunicación personal). Los presentes inventores encontraron que la expresión de *Fendrr* en heterocigóticos se restringió a los pulmones en ratones adultos (Tabla 2 en la Figura 11) y fue prominente en el tracto respiratorio en desarrollo en embriones (Fig. 2A). Puede que no de forma sorprendente, los homocigóticos *Fendrr* mostraron un estrés respiratorio y una muerte perinatal posterior debido a la maduración estructural defectuosa de los pulmones. El fenotipo nuligénico para *Fendrr* de los presentes investigadores se asemeja al trastorno raro del desarrollo pulmonar letal humano displasia alveolo-capilar con desalineación de las venas pulmonares (ACD/MPV; del inglés, human lethal lung development disorder alveolar capillary dysplasia with misalignment of pulmonary veins), en el que los pacientes muestran una deficiencia en el desarrollo de los lóbulos pulmonares y sufren una dificultad respiratoria posnatal dentro de los minutos a horas tras el nacimiento (Bishop NB, *et al.* (2011) *American Journal of Respiratory and Critical Care Medicine* 184: 172-179). Al menos se informó de un paciente de ACD/MPV que tenía una deleción de 11 kb dentro del gen *FOXF1-AS1*, el homólogo humano de *Fendrr*, expresado en pulmones de seres humanos recién nacidos (Szafranski P, *et al.* (2013)

Genome Research 23: 23-33). Grote *et al.* (Grote P, *et al.* (2013), Developmental Cell 24: 206-214) informaron de un ratón mutante con una mutación del gen *Fendrr* que produjo letalidad alrededor de E13,75 asociada con una hernia umbilical, una reducción en el grosor de la pared corporal ventral y un defecto cardíaco que provoca acumulación de sangre en el atrio derecho. Los presentes inventores no observaron ninguno de estos fenotipos. Las discrepancias entre los fenotipos se pueden explicar por los diferentes diseños alélicos. El alelo de los presentes inventores suprimió el exón 2 de *Fendrr* hasta el final, diseñado para evitar la alteración del promotor que se podría compartir con el gen codificante de proteínas *Foxf1* adyacente. El alelo *Fendrr* de Grote *et al.* (Grote P, *et al.* (2013), Developmental Cell 24: 206-214) consistió en la inserción de un elemento de parada transcripcional en el primer exón y no incluyó un gen indicador.

El patrón de expresión en adultos más destacable observado por los presentes inventores fue el de *Pint*, que exhibió un incremento en la extensión e intensidad de la tinción de X-gal a medida que los ratones envejecían desde recién nacidos a adultos maduros (Fig. 4). Este patrón llamativo asociado a la edad llevó a los presentes investigadores a realizar un análisis longitudinal para la tasa de crecimiento y signos claros de salud anormal. En comparación con los ratones WT, los presentes investigadores encontraron que a medida que los ratones *Pint*^{-/-} envejecían, en exhibieron pérdida de pelo progresivo y signos de debilitamiento muscular, cifolordosis grave, grasa corporal y densidad mineral ósea reducidas, tasa de crecimiento más lenta y supervivencia reducida. Sorprendentemente, estos resultados se replicaron en los ratones heterocigóticos, pero en menor medida. Este espectro de fenotipos asociados a la edad, junto con el incremento inusual de la expresión génica con la edad, implica que los ratones podrían requerir una dosis crítica de *Pint* para el mantenimiento general de la salud y de la función tisular durante el ciclo de vida normal y por primera vez apunta al papel potencial del ARNlinc en el envejecimiento fisiológico. Un estudio reciente mostró que *Pint* es una diana directa para p53, proporcionando un enlace entre la ruta de p53 y un silenciamiento epigenético por el complejo represor polycomb 2 (PRC2; del inglés, polycomb repressive complex 2) (Marin-Bejar *et al.*, 2013). Un número creciente de pruebas ha implicado el papel crucial de p53 en la senescencia celular y el control del envejecimiento. Será de gran interés investigar la regulación de *Pint* y su implicación potencial en la senescencia celular dependiente de p53 y el envejecimiento de los organismos. Esto podría revelar mecanismos clave en el proceso de envejecimiento fisiológico en mamíferos con indicaciones clínicas potenciales en enfermedades humanas que incluyen aquellas asociadas con envejecimiento y cáncer.

La intención de los presentes investigadores al iniciar este trabajo no fue solo arrojar luz sobre las funciones de los ARNlinc en particular cuyos genes se eligieron para mutar, sino también obtener una mejor comprensión de las propiedades generales de los ARNlinc como una clase. Esta colección podría servir como una semilla para un esfuerzo a gran escala para mutar muchos más miembros de la familia génica de ARNlinc. Se ha observado que muchos ARNlinc están asociados con proteínas que participan en la regulación de la transcripción al nivel de cromatina. Esto podría sugerir una amplia, general y redundante función en la expresión génica muy parecida a la interacción de miARN en el mantenimiento de perfiles de expresión génica específicos de tejidos a nivel postranscripcional. Los resultados de los presentes investigadores, sin embargo, parecen apuntar en una dirección diferente. Los fenotipos únicos y los patrones específicos exquisitos descritos en el presente documento y en Sauvageau *et al.* (Sauvageau M, *et al.* (2013) Elife 2: e01749) sostienen funciones específicas, directas y determinantes para los ARNlinc. Aunque este estudio es solo el comienzo del análisis de esta colección de ratones nuligénicos, revela los ARNlinc como una nueva clase de moléculas codificadas funcionales que, como las proteínas, cumplen diversos papeles en el desarrollo embrionario, fisiología y homeostasis de una amplia matriz de tejidos y órganos en mamíferos.

LISTADO DE SECUENCIAS

<110> Regeneron Pharmaceuticals, Inc.
Presidente y Miembros del Harvard College

<120> Animales no humanos deficientes en ARNlinc

<130> N407341EP-A

<150> 61/863.147

<151> 07/08/2013

<160> 2

<170> PatentIn versión 3.5

<210> 1

<211> 10

<212> ADN

<213> Secuencia artificial

<220>

<223> Secuencia consenso Kozak

ES 2 812 242 T3

5 <220>
<221> misc_feature
<222> 1, 4
<223> R es A o G

<400> 1
rccrccatgg 10

10 <210> 2
<211> 13
<212> ADN
<213> Secuencia artificial

15 <220>
<223> Secuencia consenso Kozak

20 <220>
<221> misc_feature
<222> 7
<223> R es A o G

<400> 2
gccgccrcca tgg 13

REIVINDICACIONES

1. Un ratón cuyo genoma comprende una inactivación de un locus del ARN no codificante largo (ARNlnc) *Pint* endógeno, en donde la inactivación

5

(i) da como resultado que el ratón muestra un fenotipo asociado a envejecimiento prematuro que comprende

(a) una tasa de crecimiento más lenta que la de un control de tipo silvestre;

(b) un declive en la fuerza muscular;

10

(c) fibrosis;

(d) un contenido de grasa corporal más bajo que el del control de tipo silvestre;

(e) una densidad mineral ósea del fémur y una masa ósea más baja que la del control de tipo silvestre;

(f) una masa muscular disminuida en comparación con la del control de tipo silvestre;

15

(g) una disminución en la longevidad media;

(h) cifolordosis; o

(i) una combinación de cualquiera de (a)-(h); y

(ii) comprende una delección de la mayor parte de la secuencia codificante del ARNlnc *Pint* que comienza en un segundo exón del locus del ARNlnc *Pint* endógeno y da como resultado una mutación de pérdida de función de *Pint*.

20

2. El ratón de la reivindicación 1, en donde:

(a) la inactivación comprende un reemplazo de la secuencia suprimida con un inserto de ácido nucleico; y/o

25

(b) el ratón es homocigótico para la inactivación del locus del ARNlnc *Pint* endógeno.

3. El ratón de la reivindicación 2(a), en donde el inserto de ácido nucleico comprende una primera secuencia de nucleótidos que codifica un indicador.

30

4. El ratón de la reivindicación 3, en donde:

(a) la primera secuencia de nucleótidos está ligada operativamente a un promotor que impulsa la expresión del indicador;

35

(b) la primera secuencia de nucleótidos que codifica el indicador se posiciona en un locus del ARNlnc *Pint* en unión operativa con un promotor del ARNlnc *Pint* endógeno, en donde el promotor del ARNlnc *Pint* endógeno impulsa la expresión de la primera secuencia de nucleótidos;

(c) la primera secuencia de nucleótidos comprende una secuencia consenso Kozak; o

(d) el indicador es cualquiera de β -galactosidasa, proteína fluorescente verde (GFP), proteína fluorescente verde potenciada (eGFP), mPlum, mCherry, tdTomato, mStrawberry, J-Red, DsRed, mOrange, mKO, mCitrina, Venus, YPet, proteína fluorescente amarilla potenciada (EYFP), Esmeralda, CyPet, proteína fluorescente cian (CFP), Cerulean, T-Zafiro, luciferasa, fosfatasa alcalina o una combinación de los mismos.

40

5. El ratón de la reivindicación 4(b), en donde la expresión de la primera secuencia de ácido nucleico sigue un patrón de expresión del ARNlnc *Pint*.

45

6. El ratón de la reivindicación 2(a), en donde el inserto de ácido nucleico comprende una segunda secuencia de ácido nucleico que codifica un marcador seleccionable, en donde la segunda secuencia de ácido nucleico está ligada operativamente a un promotor, opcionalmente en donde el inserto de ácido nucleico comprende sitios de recombinación específica de sitio que flanquean un segmento que codifica el indicador y/o un segmento que codifica el marcador seleccionable.

50

7. El ratón de la reivindicación 1, en donde la inactivación comprende un reemplazo de la secuencia suprimida con un inserto de ácido nucleico,

en donde el inserto de ácido nucleico comprende una secuencia nucleotídica que codifica para uno o los dos de (i) β -galactosidasa y (ii) un marcador seleccionable y

55

en donde el inserto de ácido nucleico se posiciona en el locus del ARNlnc *Pint* en unión operativa con un promotor del ARNlnc *Pint* endógeno que impulsa la expresión del inserto de ácido nucleico.

8. El ratón de la reivindicación 7, en donde:

60

(a) el inserto de ácido nucleico reemplaza al menos 30 kb del locus *Pint* endógeno; o

(b) el inserto de ácido nucleico reemplaza la secuencia genómica entre las coordenadas genómicas del Cromosoma 6: 31166026-31197846.

65

9. El ratón de la reivindicación 8(b), en donde el inserto de ácido nucleico comprende adicionalmente una secuencia consenso Kozak.

10. El ratón de la reivindicación 9, en donde:

- 5 (a) el inserto de ácido nucleico codifica β galactosidasa pero no el marcador seleccionable;
 (b) el inserto de ácido nucleico codifica el marcador seleccionable pero no β galactosidasa;
 (c) el inserto de ácido nucleico comprende sitios de recombinación específica de sitio que flanquean un segmento que codifica β galactosidasa y/o un segmento que codifica el marcador seleccionable; o
 (d) el ratón es homocigótico para el locus del ARNinc *Pint* endógeno modificado.

10 11. Una célula de ratón, tejido o embrión cuyo genoma comprende una inactivación de un locus del ARN no codificante largo (ARNlnc) *Pint* endógeno, en donde la inactivación comprende una delección de la mayor parte de la secuencia codificante para el ARNlnc *Pint* que comienza en un segundo exón del locus del ARNlnc *Pint* endógeno y da como resultado una mutación de pérdida de función de *Pint*.

15 12. La célula, tejido o embrión de la reivindicación 11, en donde la inactivación da como resultado un fenotipo asociado a envejecimiento prematuro que comprende

- (a) una tasa de crecimiento más lenta que la de un control de tipo silvestre;
 (b) un declive en la fuerza muscular;
 20 (c) fibrosis;
 (d) un contenido de grasa corporal más bajo que el del control de tipo silvestre;
 (e) una densidad mineral ósea del fémur y una masa ósea más baja que la del control de tipo silvestre;
 (f) una masa muscular disminuida en comparación con la del control de tipo silvestre;
 (g) una disminución en la longevidad media;
 25 (h) cifolordosis; o
 (i) una combinación de cualquiera de (a)-(h)

en un ratón cuyo genoma comprende la inactivación.

30 13. La célula, tejido o embrión de la reivindicación 11 o 12, en donde la inactivación comprende un reemplazo de al menos un kilobase del locus del ARNlnc *Pint* con un inserto de ácido nucleico, en donde el reemplazo comienza en el segundo exón del locus del ARNlnc *Pint* y da como resultado una pérdida de función del locus del ARNlnc *Pint*, en donde el inserto de ácido nucleico comprende una secuencia nucleotídica que codifica para uno o los dos de (i) β -galactosidasa y (ii) un marcador seleccionable y
 35 en donde el inserto de ácido nucleico se posiciona en el locus del ARNlnc *Pint* en unión operativa con un promotor del ARNlnc *Pint* endógeno que impulsa la expresión del inserto de ácido nucleico.

14. La célula de una cualquiera de las reivindicaciones 11 a 13, en donde la célula es una célula madre embrionaria.

40 15. Un método para hacer el ratón de una cualquiera de las reivindicaciones 1 a 10 que comprende:

- (1) poner en contacto una célula madre embrionaria (ES) de ratón con un constructo de direccionamiento que comprende un inserto de ácido nucleico flanqueado por brazos de homología 5' y 3'; en donde el constructo de direccionamiento sufre recombinación homóloga con el locus del ARNlnc *Pint* endógeno de ratón en un genoma
 45 de la célula ES de ratón para formar una célula ES de ratón modificada;
 (2) introducir la célula ES de ratón modificada en un embrión de ratón hospedador; y
 (3) gestar el embrión de ratón hospedador en una madre ratón subrogada, en donde la madre ratón subrogada produce una prole de ratones que comprende la inactivación del locus del ARNlnc *Pint* endógeno.

50 16. El método de la reivindicación 15, en donde poner en contacto la célula ES de ratón comprende reemplazar la secuencia suprimida dentro del locus del ARNlnc *Pint* comenzando en el segundo exón del locus del ARNlnc *Pint* con el inserto de ácido nucleico.

Figura 1

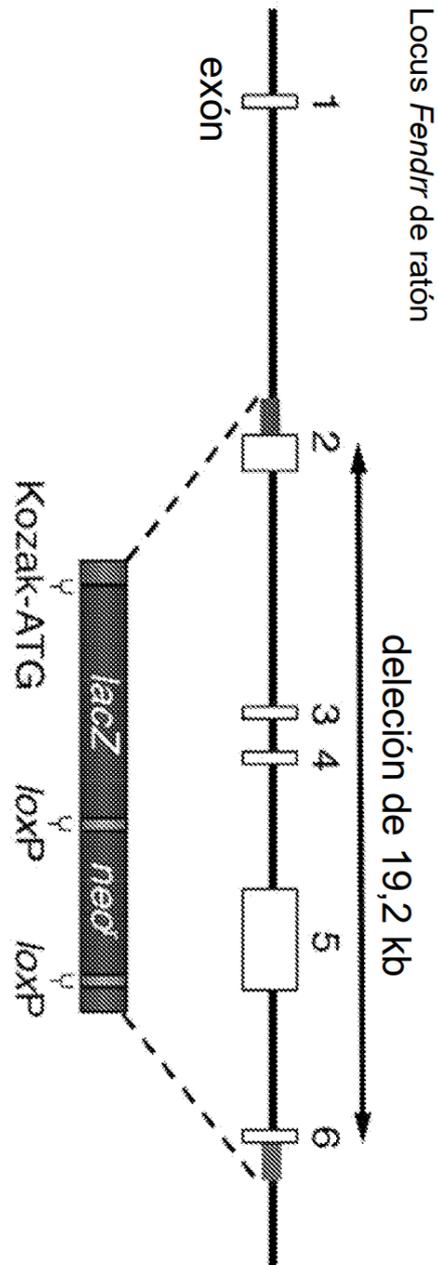


Figura 2A

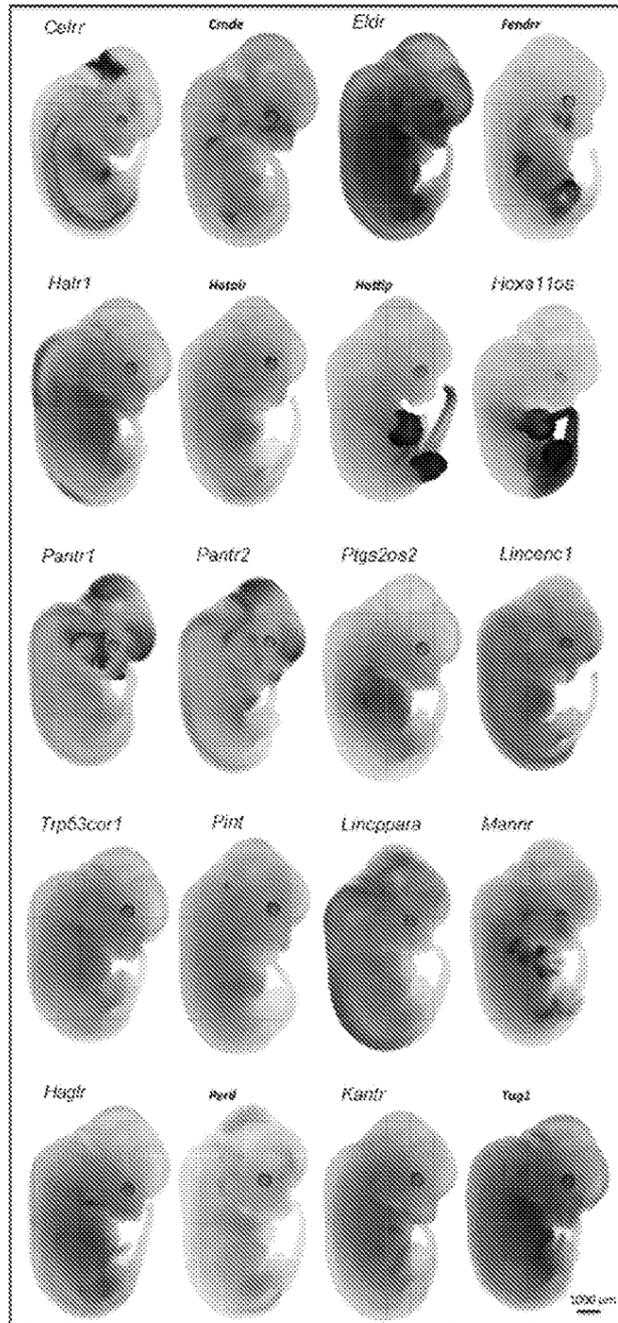


Figura 2B

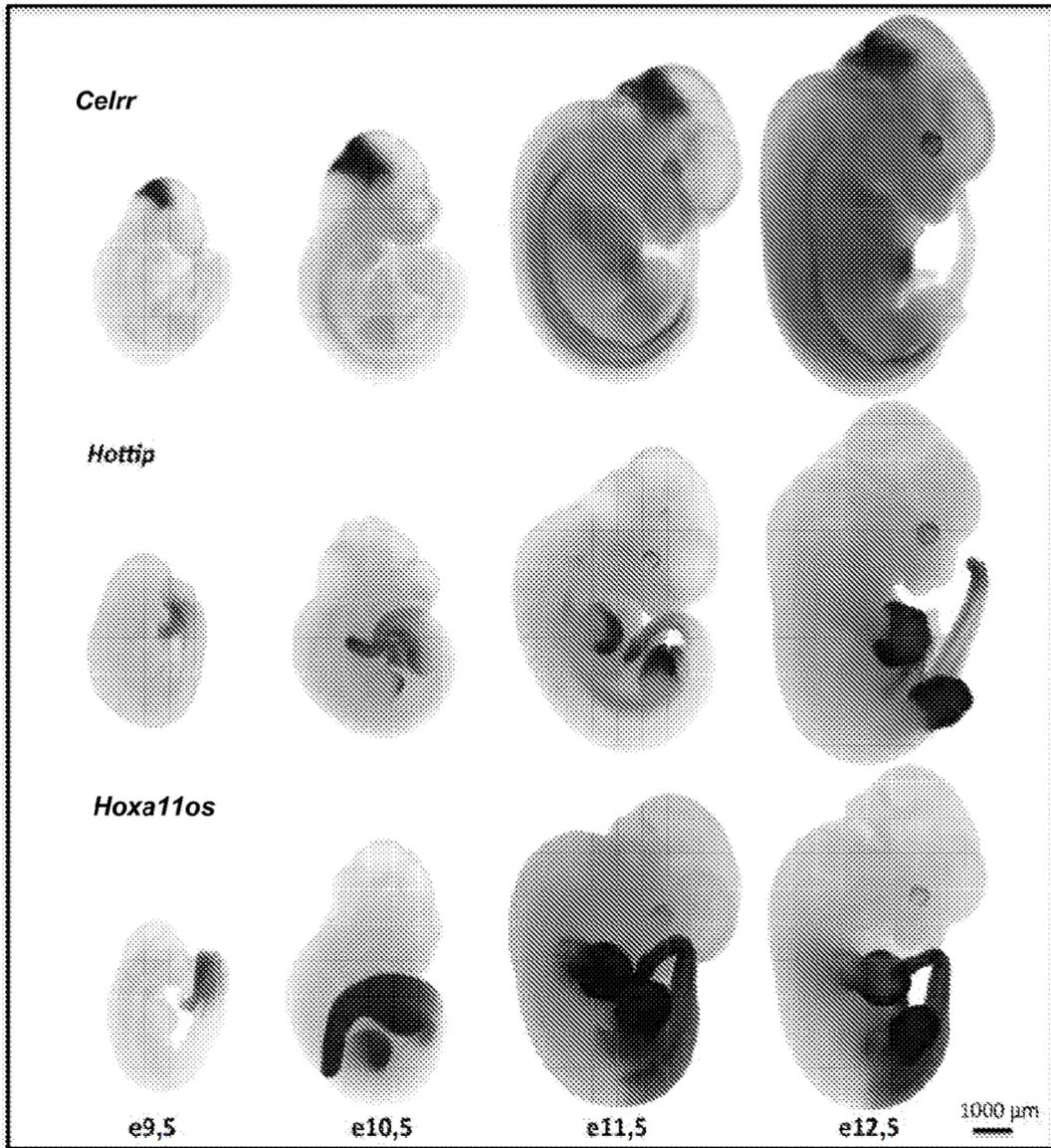


Figura 3

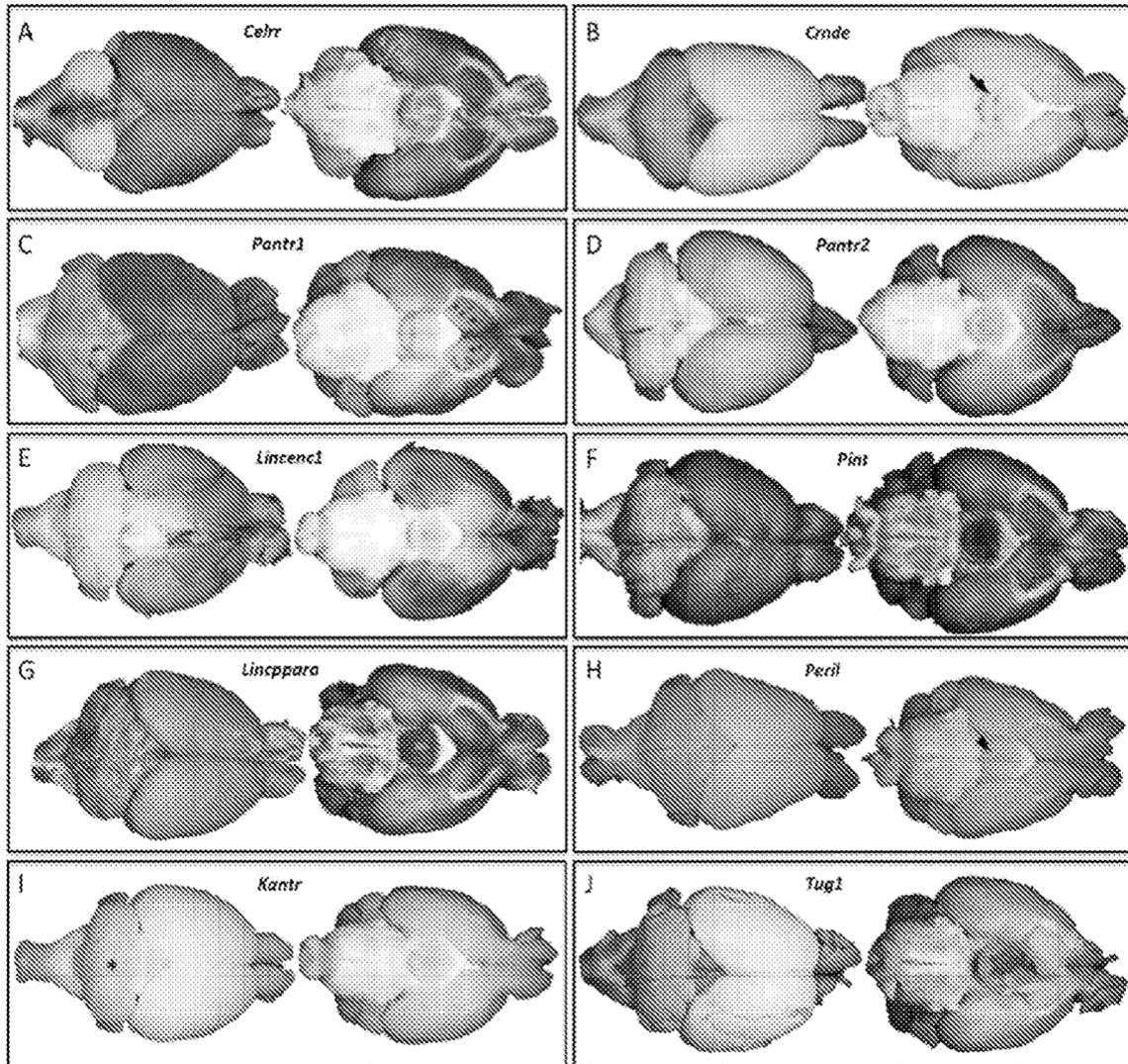


Figura 4

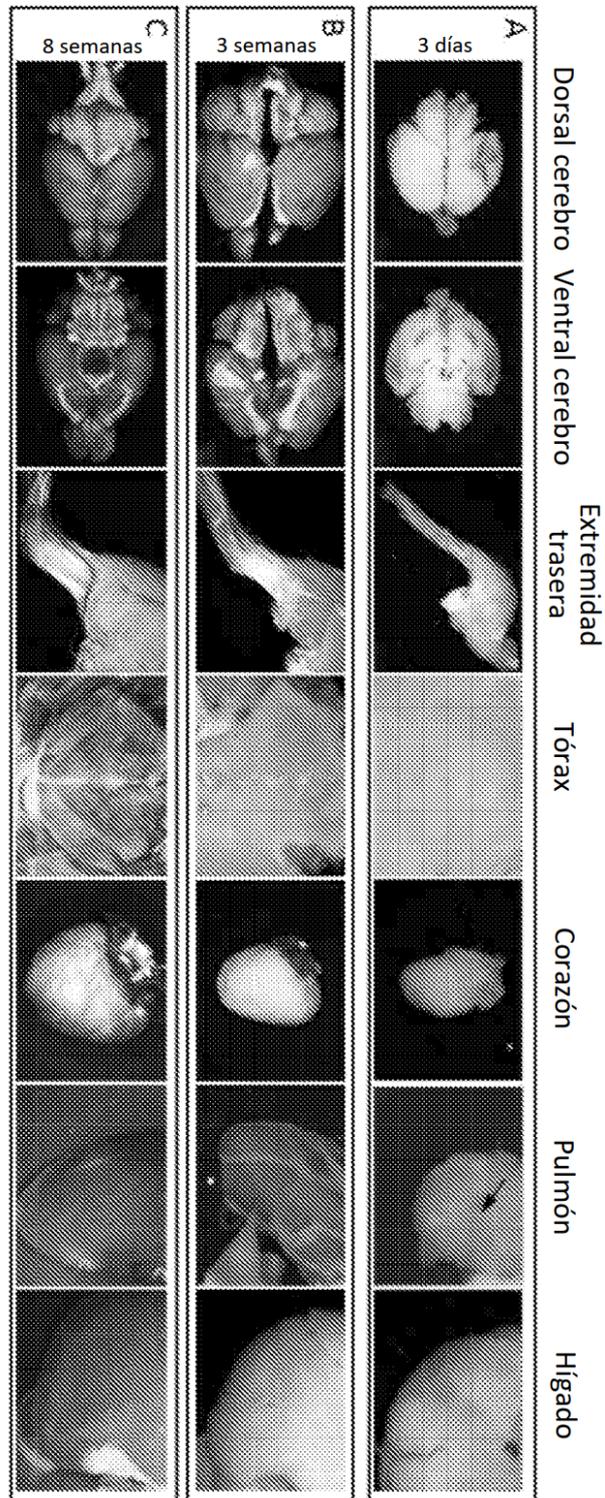


Figura 5

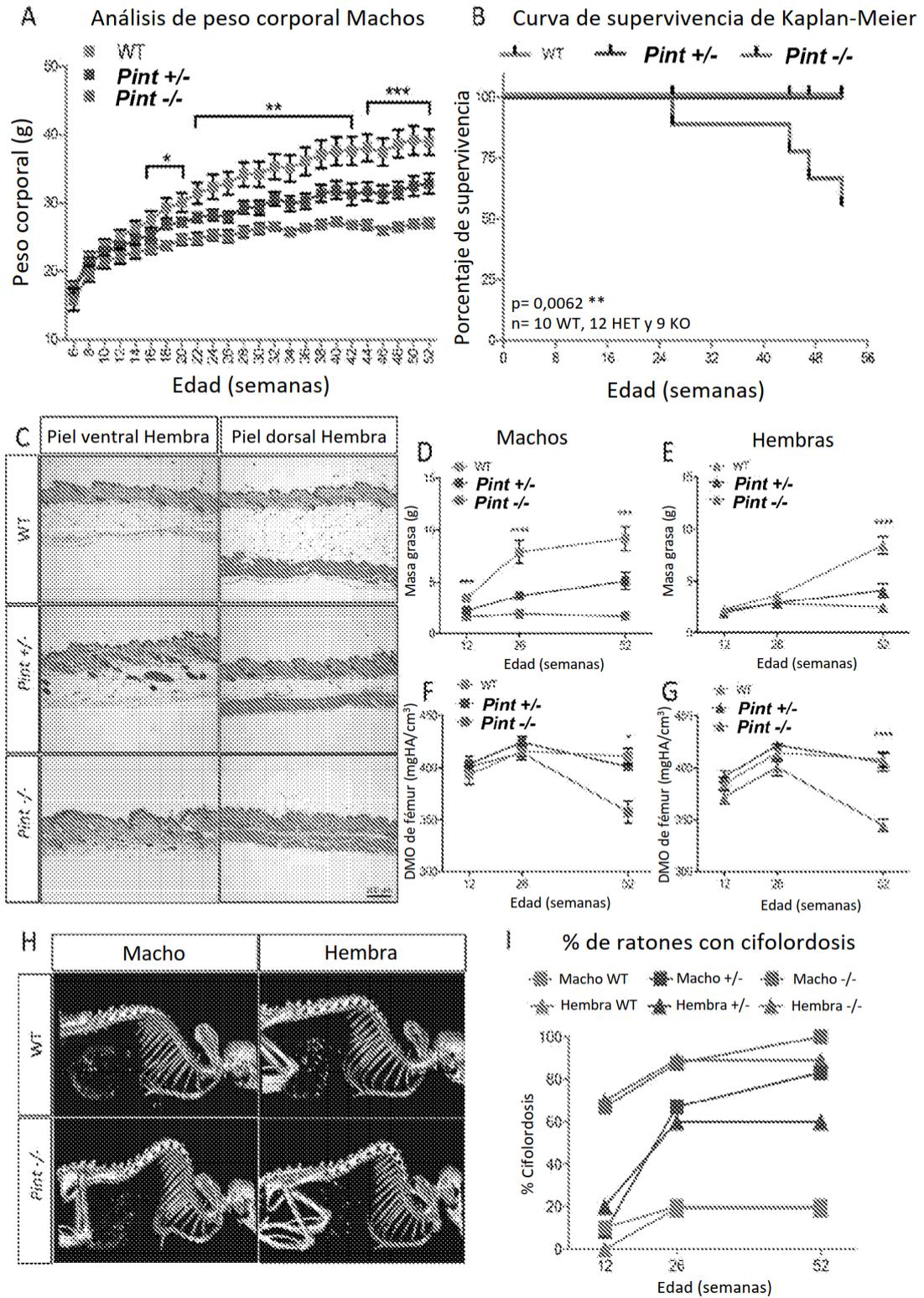


Figura 6

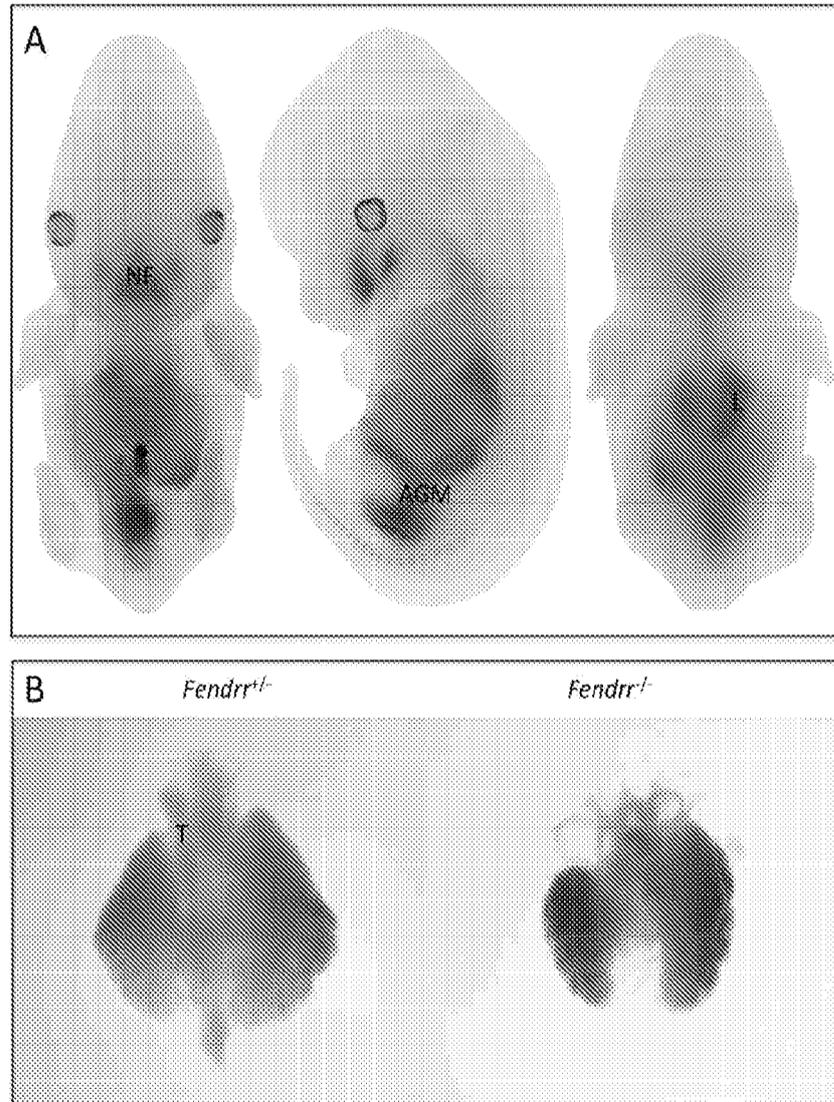


Figura 7

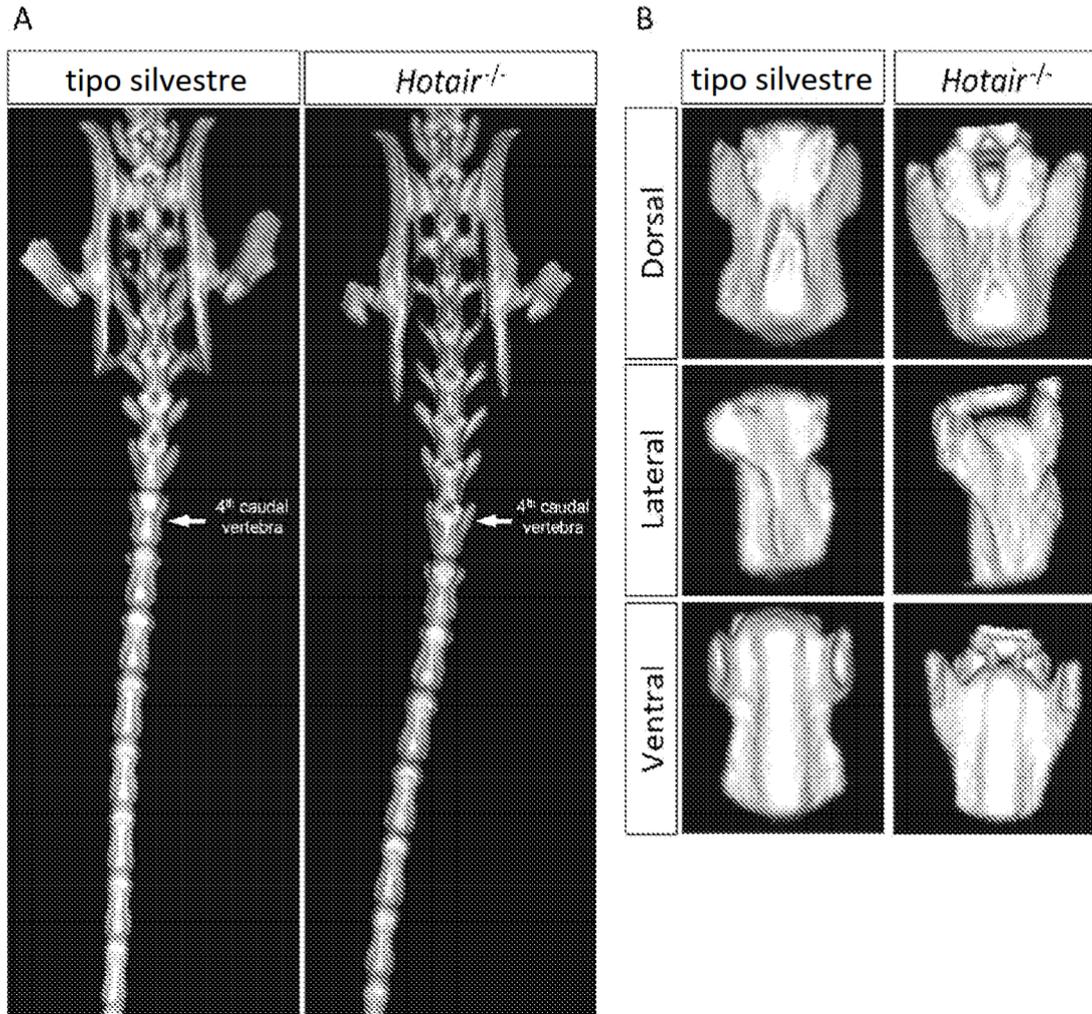


Figura 8

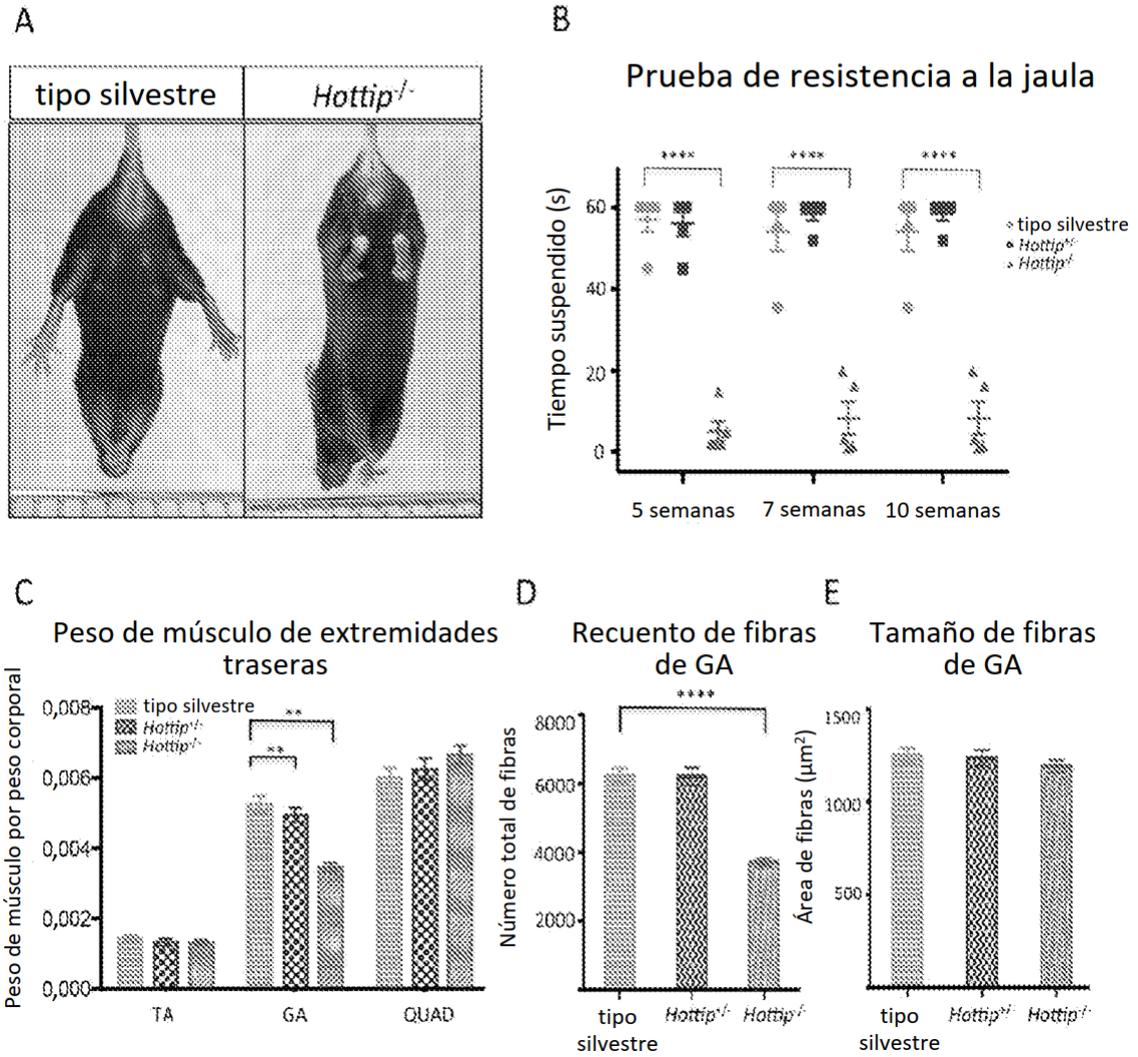


Figura 9

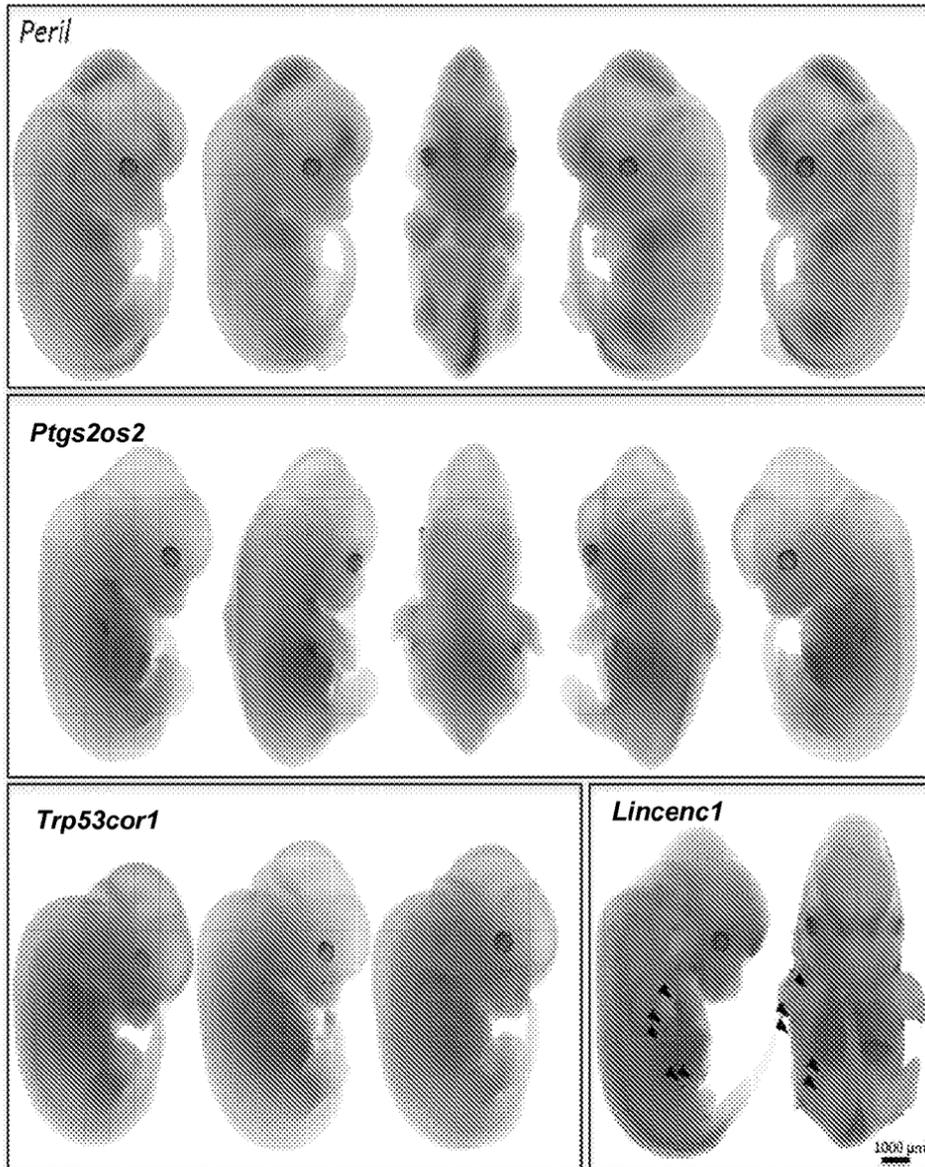


Figura 10

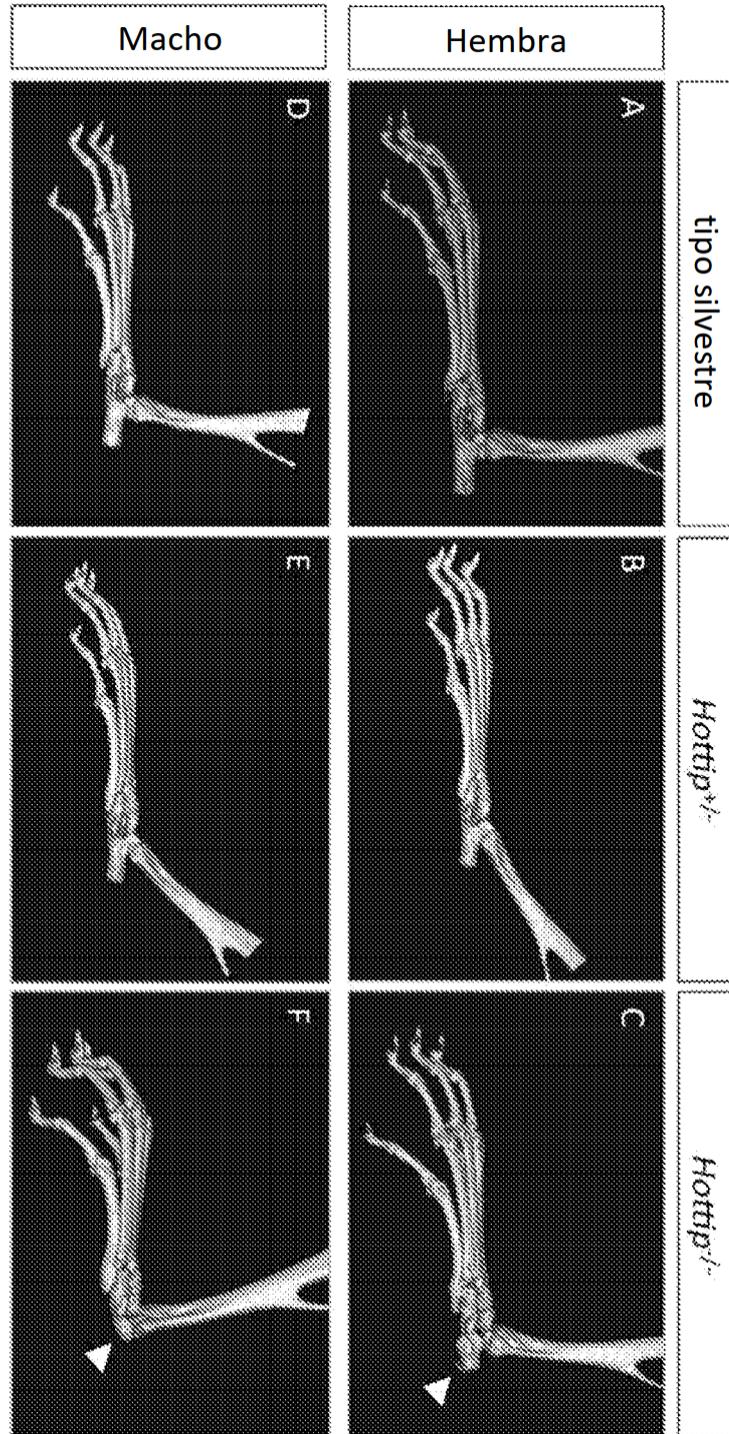


Figura 11

Tabla 2. Sumario de la expresión del indicador de *lacZ* en embrión y tejido de adultos

ARNlinec	Embrión	Cerebro	Corazón	Pulmones	Hígado	Costillas	Bazo	Intestino	Estómago	Riñón	Urogenital	Extremidad trasera
<i>Celr</i>	+	+				+						+
<i>Cmde</i>	+	+									+	+
<i>Eldr</i>	+										+	
<i>Fendr</i>	+			+								
<i>Hair1</i>	+					+						
<i>Holeir</i>	+											
<i>Hoftip</i>	+							+			+	+
<i>Hoxa11os</i>	+					+		+		+	+	+
<i>Pantr1</i>	+	+				+				+		
<i>Pantr2</i>	+	+										
<i>Ptgs2os2</i>	+											
<i>incent1</i>	+	+										+
<i>Ttp53cor1</i>	+									+	+	
<i>Pyt</i>	+	+		+		+		+		+	+	+
<i>lncppara</i>	+		+		+				+			+
<i>Manntr</i>	+			+								
<i>Haglr</i>	+											
<i>Peril</i>	+	+										
<i>Kantr</i>	+	+										
<i>Tug1</i>	+	+			+					+	+	