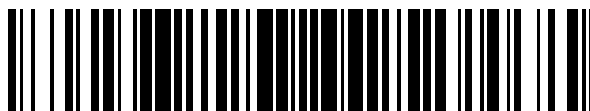


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 777 894**

51 Int. Cl.:

G16B 5/20 (2009.01)

G16B 25/10 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.03.2012 PCT/US2012/027615**

87 Fecha y número de publicación internacional: **07.09.2012 WO12119129**

96 Fecha de presentación y número de la solicitud europea: **02.03.2012 E 12751945 (2)**

97 Fecha y número de publicación de la concesión europea: **20.11.2019 EP 2680925**

54 Título: **Ensayos por interrogación basados en células y usos de los mismos**

30 Prioridad:

02.03.2011 US 201161448587 P
01.02.2012 US 201261593848 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.08.2020

73 Titular/es:

BERG LLC (100.0%)
500 Old Connecticut Path, Building B
Framingham, MA 01701, US

72 Inventor/es:

NARAIN, NIVEN, RAJIN;
SARANGARAJAN, RANGAPRASAD y
VISHNUDAS, VIVEK, K.

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 777 894 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

Ensayos por interrogación basados en células y usos de los mismos

5 **Referencia cruzada a solicitudes relacionadas**

Esta solicitud reivindica la prioridad de la solicitud provisional estadounidense con n.º de serie 61/448.587, presentada el 2 de marzo de 2011 y la solicitud provisional estadounidense con n.º de serie 61/593.848, presentada el 2 de febrero de 2012.

10

Antecedentes de la invención

El desarrollo de nuevos fármacos ha mejorado mucho desde el descubrimiento del ADN en 1964 por James Watson y Francis Crick, pioneros de lo que se denomina hoy en día biología molecular. Las herramientas y los productos de la biología molecular permiten una medición rápida, detallada y precisa de la regulación génica tanto a nivel de ADN como de ARN. Las siguientes tres décadas tras el descubrimiento del cambio de paradigma vieron la génesis de modelos animales deficientes, reacciones clave relacionadas con enzimas y una comprensión nueva de los mecanismos de enfermedad y la fisiopatología de las plataformas mencionadas anteriormente. En la primavera de 2000, cuando Craig Venter y Francis Collins anunciaron la secuencia inicial del genoma humano, el mundo científico entró en una nueva corriente de la medicina.

15

20

25

El mapeo del genoma despertó inmediatamente la esperanza de, por ejemplo, poder controlar la enfermedad incluso antes de que se iniciara, de usar la terapia génica para revertir los procesos cerebrales degenerativos que provocan la enfermedad de Alzheimer o Parkinson, y de un constructo que pudiera introducirse en un sitio tumoral y provocar la erradicación de la enfermedad mientras se restauraba la arquitectura y fisiología del tejido normal. Otros tomaron giros controvertidos y propusieron la noción de crear la descendencia deseada con respecto al color de los ojos o el cabello, la estatura, etc. Sin embargo, diez años después todavía se está esperando sin ningún camino particular a la vista para el éxito sostenido de la terapia génica, o incluso el control elemental del proceso genético.

30

Por tanto, una realidad aparente es que la genética, al menos independientemente de los constructos de soporte, no controla el punto final de la fisiología. De hecho, muchos procesos tales como modificaciones postranscripcionales, las mutaciones, los polimorfismos de un solo nucleótido (SNP) y las modificaciones traduccionales podrían alterar la providencia de un gen y/o su proteína complementaria codificada, y contribuir de ese modo al proceso de enfermedad.

35

Se ha hallado que el documento US 008/208784 A1 está relacionado con la presente invención. El documento US 008/208784 A1 divulga sistemas y métodos que utilizan un marco de modelado probabilístico para la ingeniería inversa de un conjunto de modelos causales, a partir de datos y luego sometiendo a simulación hacia delante el conjunto de modelos para analizar y predecir el comportamiento de la red.

40

Además, se ha hallado que el documento GILL RYAN *ET AL*: "A statistical framework for differential network analysis from microarray data", BMC BIOINFORMATICS, BIOMED CENTRAL, LONDRES, GB, vol. 11, n.º 1, 19 de febrero de 2010, página 95, XP021065695, ISSN: 1471-2105 está relacionado con la presente invención y divulga la realización de un análisis diferencial de redes construidas a partir de datos de microalineamientos en dos entornos experimentales.

45

Sumario de la invención

La era de la información y la creación de Internet han permitido una sobrecarga de información, al tiempo que se facilita la colaboración y la crítica internacional. Irónicamente, las realidades mencionadas anteriormente también pueden ser la causa de que la comunidad científica pase por alto algunos puntos simples, incluyendo que la comunicación de cascadas de señales y la interferencia dentro y entre células y/o tejidos permite que se produzcan homeostasis y mensajes para mecanismos correctores cuando algo va mal.

50

Un caso específico se refiere a la enfermedad cardiovascular (ECV), que sigue siendo la principal causa de muerte en los Estados Unidos y gran parte del mundo desarrollado, representando 1 de cada 2,8 muertes sólo en los Estados Unidos. Además, la ECV sirve como patología subyacente que contribuye a las complicaciones asociadas, tales como enfermedad renal crónica (~ 19 millones de casos en EE.UU.), síndrome de fatiga crónica, y un factor clave en el síndrome metabólico. Los avances significativos en la tecnología relacionada con el diagnóstico, las técnicas quirúrgicas mínimamente invasivas, los stents farmacoactivos y la vigilancia clínica efectiva han contribuido a un periodo de crecimiento incomparable en el campo de la cardiología intervencionista y han permitido un manejo más efectivo de la ECV. Sin embargo, la etiología de la enfermedad relacionada con ECV y las comorbilidades asociadas tales como diabetes y vasculopatía periférica aún no se han dilucidado por completo.

55

60

Todavía se carece de nuevos enfoques para explorar los mecanismos y las rutas implicadas en un proceso biológico, tales como la etiología de los estados de enfermedad (por ejemplo, ECV), y para identificar rutas

65

reguladoras clave y/o moléculas diana (por ejemplo, “dianas con capacidad farmacológica”) y/o marcadores para un mejor diagnóstico, manejo y/o tratamiento de la enfermedad.

5 La invención descrita en el presente documento se basa, al menos en parte, en una utilización nueva y colaborativa de herramientas y metodologías de red de biología, genómica, proteómica, metabólica, transcriptómica y bioinformática que, cuando se combinan, pueden usarse para estudiar cualquier sistema biológico de interés, tales como estados de enfermedad seleccionados incluyendo cáncer, diabetes, obesidad y enfermedad cardiovascular, usando un enfoque de biología de sistemas. En una primera etapa, se desarrollan sistemas de modelado celular para examinar con sonda diversos sistemas biológicos, tales como un proceso de enfermedad, que comprende 10 células relacionadas con la enfermedad sujetas a diversos estímulos ambientales relevantes para la enfermedad (por ejemplo, hiperglucemia, hipoxia, estrés inmunológico y peroxidación lipídica). En algunas realizaciones, el sistema de modelado celular implica mecanismos de interferencia celular entre diversos tipos de células que interactúan (tales como células de músculo liso aórtico (HASMC), células renales del túbulo proximal (HK-2), células endoteliales aórticas (HAEC) y fibroblastos dérmicos (HDFa)). Se obtienen lecturas biológicas de alto rendimiento del sistema de modelo celular mediante el uso de una combinación de técnicas incluyendo, por ejemplo, 15 espectrometría de masas de vanguardia (CL/EM/EM), citometría de flujo, ensayos basados en células y ensayos funcionales. Las lecturas biológicas de alto rendimiento se someten entonces a un análisis bioinformático para estudiar las tendencias de datos congruentes mediante modelado *in vitro*, *in vivo* e *in silico*. Los alineamientos resultantes permiten la minería de datos con relaciones cruzadas, donde se desarrollaron análisis de regresión lineal y no lineal para alcanzar puntos de presión concluyentes (o “centros”). Estos “centros”, tal como se presentan en el presente documento, son candidatos para el descubrimiento de fármacos. En particular, estos centros representan posibles dianas farmacológicas y/o marcadores de enfermedad.

25 Las firmas moleculares de los diferenciales permiten conocer los mecanismos que dictan las alteraciones en el microentorno tisular que conducen a la aparición y progresión de la enfermedad. Tomados en conjunto, la combinación de las plataformas tecnológicas mencionadas anteriormente con el modelado celular estratégico permite una inteligencia robusta que puede emplearse para establecer adicionalmente la comprensión de la enfermedad al tiempo que se crean bibliotecas de biomarcadores y candidatos a fármacos que pueden mejorar clínicamente las normas asistenciales en cardiología intervencionista.

30 Además, este enfoque no sólo es útil para el diagnóstico de o la intervención en la enfermedad, sino que también tiene aplicabilidad general a prácticamente todos los estados patológicos o no patológicos en los sistemas biológicos, tales como sistemas biológicos en los que interactúan dos o más sistemas celulares. Por ejemplo, este enfoque es útil para obtener información sobre los mecanismos asociados o causales de toxicidad farmacológica. Por tanto, la invención proporciona un marco para una evaluación biológica por interrogación que puede aplicarse 35 generalmente en un amplio espectro de configuraciones.

Una característica importante de la plataforma de la invención es que el sistema basado en IA se basa en los grupos de datos obtenidos del sistema de modelo celular, sin recurrir ni tener en cuenta ningún conocimiento existente en la técnica, tal como relaciones biológicas conocidas (es decir, no hay puntos de datos artificiales), referentes al proceso biológico. Por consiguiente, los modelos estadísticos resultantes generados a partir de la plataforma no tienen sesgo. Otra característica importante de la plataforma de la invención y sus componentes, por ejemplo, los sistemas de modelo celular y los grupos de datos obtenidos a partir de los mismos, es que permite la construcción continua en los modelos celulares a lo largo del tiempo (por ejemplo, mediante la introducción de nuevas células y/o 45 condiciones), de modo que una red de relaciones causales de consenso inicial, de “primera generación”, generada a partir de un modelo celular para un sistema o proceso biológico puede evolucionar junto con la evolución del propio modelo celular a una red de relaciones causales de generación múltiple (y redes delta o delta-delta obtenidas a partir de la misma). De esta manera, tanto los modelos celulares, los grupos de datos procedente de los modelos celulares y las redes de relaciones causales generadas a partir de los modelos celulares usando los métodos de la tecnología de plataforma pueden evolucionar constantemente y sumarse al conocimiento previo obtenido de la tecnología de 50 plataforma.

Los problemas mencionados se resuelven mediante el contenido de las reivindicaciones independientes. Otras realizaciones preferidas se definen en las reivindicaciones dependientes.

55 En determinadas realizaciones, la “perturbación ambiental”, también denominada en el presente documento “componente de estímulo externo”, es un agente terapéutico. En determinadas realizaciones, el componente de estímulo externo es una molécula pequeña (por ejemplo, una molécula pequeña de no más de 5 kDa, 4 kDa, 3 kDa, 2 kDa, 1 kDa, 500 Dalton o 250 Dalton). En determinadas realizaciones, el componente de estímulo externo es un producto biológico. En determinadas realizaciones, el componente de estímulo externo es un producto químico. En determinadas realizaciones, el componente de estímulo externo es endógeno o exógeno con respecto a las células. En determinadas realizaciones, el componente de estímulo externo es Una MIM o intercambiador epimetabólico (*epishifter*). En determinadas realizaciones, el componente de estímulo externo es un factor de estrés para el sistema celular, tal como hipoxia, hiperglucemia, hiperlipidemia, hiperinsulinemia y/o condiciones ricas en ácido 65 láctico.

En determinadas realizaciones, el componente de estímulo externo puede incluir un agente terapéutico o un agente terapéutico candidato para tratar un estado de enfermedad, incluyendo agente quimioterapéutico, fármacos biológicos basados en proteínas, anticuerpos, proteínas de fusión, fármacos de molécula pequeña, lípidos, polisacáridos, ácidos nucleicos, etc.

5 En determinadas realizaciones, el componente de estímulo externo puede ser uno o más factores de estrés, tales como los que se encuentran normalmente *in vivo* en los diversos estados de enfermedad, incluyendo hipoxia, condiciones hiperglucémicas, entorno ácido (que puede imitarse mediante tratamiento con ácido láctico), etc.

10 En otras realizaciones, el componente de estímulo externo puede incluir uno o más MIM y/o intercambiadores epimetabólicos, tal como se definen a continuación en el presente documento. Las MIM a modo de ejemplo incluyen la coenzima Q10 (también denominada en el presente documento CoQ10) y compuestos en la familia de la vitamina B, o nucleósidos, mononucleótidos o dinucleótidos que comprenden un compuesto en la familia de la vitamina B.

15 Al realizar mediciones de salida celular (tal como expresión de proteína), puede usarse o bien una cantidad absoluta (por ejemplo, cantidad de expresión) o bien un nivel relativo (por ejemplo, nivel de expresión relativo). En una realización, se usan cantidades absolutas (por ejemplo, cantidades de expresión). En una realización, se usan las cantidades o los niveles relativos (por ejemplo, niveles relativos de expresión). Por ejemplo, para determinar el nivel de expresión relativo de proteína de un sistema celular, la cantidad de cualquier proteína dada en el sistema celular, con o sin el estímulo externo para el sistema celular, puede compararse con una línea celular de control adecuada o una mezcla de líneas celulares (tales como todas las células usadas en el mismo experimento) y dado un valor de aumento en veces o disminución en veces. El experto apreciará que pueden emplearse cantidades absolutas o cantidades relativas en cualquier medición de salida celular, tal como nivel de transcripción de genes y/o ARN, nivel de lípidos o cualquier salida funcional, por ejemplo, nivel de apoptosis, nivel de toxicidad, o TAEC o TCO tal como se describen en el presente documento. Puede usarse un nivel umbral predeterminado para el aumento en veces (por ejemplo, un aumento de al menos 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 ó 100 veces o más) o disminución en veces (por ejemplo, al menos una disminución de hasta 0,9, 0,8, 0,75, 0,7, 0,6, 0,5, 0,45, 0,4, 0,35, 0,3, 0,25, 0,2, 0,15, 0,1 ó 0,05 veces, o una disminución de hasta el 90%, el 80%, el 75%, el 70%, el 65%, el 60%, el 55%, el 50%, el 45%, el 40%, el 35%, el 30%, el 25%, el 20%, el 15%, el 10% o el 5% o menos) para seleccionar diferenciales significativos, y los datos de salida celular para los diferenciales significativos pueden incluirse entonces en los grupos de datos (por ejemplo, los grupos de datos primero y segundo) utilizados en los métodos de tecnología de plataforma de la invención. Se pretende que todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, por ejemplo, entre 1,5 y 5 veces, 5 y 10 veces, 2 y 5 veces, o entre 0,9 y 0,7, 0,9 y 0,5, o 0,7 y 0,3 veces, formen parte de esta invención.

35 En la totalidad de la presente solicitud, todos los valores presentados en una lista, por ejemplo, tales como los anteriores, también pueden ser el límite superior o inferior de intervalos que se pretende que formen parte de esta invención.

40 En una realización de los métodos de la invención, no todas las relaciones causales observadas en una red de relaciones causales pueden ser de significación biológica. Con respecto a cualquier sistema biológico dado para el que se aplica la evaluación biológica por interrogación del sujeto, algunas (o tal vez todas) de las relaciones causales (y los genes asociados con las mismas) pueden ser "determinantes" con respecto al problema biológico específico en cuestión, por ejemplo, o bien responsable de provocar una enfermedad (una posible diana para intervención terapéutica) o bien es un biomarcador para el estado de enfermedad (un posible factor de diagnóstico o pronóstico). En una realización, una relación causal observada única en el sistema biológico es determinante con respecto al problema biológico específico en cuestión. En una realización, no todas las relaciones causales observadas únicas en el sistema biológico son determinantes con respecto al problema específico en cuestión.

50 Tales relaciones causales determinantes puede seleccionarlas un usuario final del método objeto, o puede seleccionarlas un programa de software de bioinformática, tal como REFS, el programa de análisis de rutas comparativo habilitado por DAVID o el programa de análisis de rutas KEGG. En determinadas realizaciones, se usa más de un programa de software de bioinformática, y se prefieren los resultados de consenso de dos o más programas de software de bioinformática.

55 Tal como se usa en el presente documento, los "diferenciales" de las salidas celulares incluyen diferencias (por ejemplo, niveles aumentados o disminuidos) en uno cualquiera o más parámetros de las salidas celulares. En determinadas realizaciones, los diferenciales se seleccionan cada uno independientemente del grupo que consiste en diferenciales en la transcripción de ARNm, expresión de proteína, actividad de proteína, nivel de producto intermedio/metabolito y/o interacción ligando-diana. Por ejemplo, en cuanto al nivel de expresión de proteína, los diferenciales entre dos salidas celulares, tales como salidas asociadas con un sistema celular antes y después del tratamiento con un componente de estímulo externo, pueden medirse y cuantificarse mediante el uso de tecnologías reconocidas en la técnica, tales como ensayos basados en espectrometría de masas (por ejemplo, iTRAQ, 2D-CL-EM/EM, etc.).

60

65

- 5 En un aspecto, el modelo celular para un sistema biológico comprende un sistema de interferencia celular, en el que un primer sistema celular que tiene un primer entorno celular con un componente de estímulo externo genera un primer entorno celular modificado; de tal manera que se establece un sistema celular de interferencia exponiendo un segundo sistema celular que tiene un segundo entorno celular al primer entorno celular modificado.
- 10 En una realización, se genera al menos un diferencial de interferencia celular significativo a partir del sistema celular de interferencia; y se identifica al menos un diferencial de interferencia celular determinante de tal manera que se produce una evaluación biológica por interrogación. En determinadas realizaciones, el al menos un diferencial de interferencia celular significativo es una pluralidad de diferenciales.
- 15 En determinadas realizaciones, el al menos un diferencial de interferencia celular determinante lo selecciona el usuario final. Alternativamente, en otra realización, el al menos un diferencial de interferencia celular determinante lo selecciona un programa de software de bioinformática (tal como, por ejemplo, REFS, análisis de rutas KEGG o análisis de rutas comparativo habilitado por DAVID) basándose en datos de proteómica cuantitativos.
- 20 En determinadas realizaciones, el método comprende además generar un diferencial de salida celular significativo para el primer sistema celular.
- 25 En determinadas realizaciones, los diferenciales se seleccionan cada uno independientemente del grupo que consiste en diferenciales en la transcripción de ARNm, expresión de proteína, actividad de proteína, nivel de producto intermedio/metabolito y/o interacción ligando-diana.
- En determinados ejemplos que respaldan la comprensión, el primer sistema celular y el segundo sistema celular se seleccionan independientemente de: una población homogénea de células primarias, una línea celular de cáncer o una línea celular normal.
- 30 En determinados ejemplos que respaldan la comprensión, el primer entorno celular modificado comprende factores secretados por el primer sistema celular en el primer entorno celular, como resultado de poner en contacto el primer sistema celular con el componente de estímulo externo. Los factores pueden comprender proteínas secretadas u otras moléculas de señalización. En determinados ejemplos que respaldan la comprensión, el primer entorno celular modificado está sustancialmente libre del componente de estímulo externo original.
- 35 En determinados ejemplos que respaldan la comprensión, el sistema celular de interferencia comprende una placa Transwell que tiene un compartimento de inserto y un compartimento de pocillo separados por una membrana. Por ejemplo, el primer sistema celular puede crecer en el compartimento de inserto (o el compartimento de pocillo), y el segundo sistema celular puede crecer en el compartimento de pocillo (o el compartimento de inserto).
- 40 En determinados ejemplos que respaldan la comprensión, el sistema celular de interferencia comprende un primer cultivo para hacer crecer el primer sistema celular y un segundo cultivo para hacer crecer el segundo sistema celular. En este caso, el primer entorno celular modificado puede ser un medio condicionado del primer sistema celular.
- 45 En determinados ejemplos que respaldan la comprensión, el primer entorno celular y el segundo entorno celular pueden ser idénticos. En determinados ejemplos que respaldan la comprensión, el primer entorno celular y el segundo entorno celular pueden ser diferentes.
- 50 En determinados ejemplos que respaldan la comprensión, el sistema celular de interferencia comprende un cocultivo del primer sistema celular y el segundo sistema celular.
- 55 Los métodos de la invención pueden usarse para, o aplicarse a, cualquier número de “evaluaciones biológicas por interrogación”. La aplicación de los métodos de la invención a una evaluación biológica por interrogación permite la identificación de uno o más moduladores de un sistema biológico o “controladores” de procesos celulares determinantes de un sistema o proceso biológico.
- 60 Los métodos de la invención pueden usarse para llevar a cabo una amplia gama de evaluaciones biológicas por interrogación. En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es el diagnóstico de un estado de enfermedad. En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la determinación de la eficacia de un fármaco. En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la determinación de la toxicidad de un fármaco. En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la estadificación de un estado de enfermedad. En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación identifica dianas para cosméticos antienvjecimiento.
- 65 Tal como se usa en el presente documento, una “evaluación biológica por interrogación” puede incluir la identificación de uno o más moduladores de un sistema biológico, por ejemplo, “controladores” de proceso celular determinantes (por ejemplo, un aumento o disminución de la actividad de una ruta biológica, o miembros clave de la

5 ruta, o reguladores clave para miembros de la ruta) asociados con la perturbación ambiental o el componente de estímulo externo, o una única relación causal única en un sistema biológico o proceso. Puede incluir además etapas adicionales diseñadas para someter a prueba o verificar si los controladores de proceso celular determinantes identificados son necesarios y/o suficientes para los eventos posteriores asociados con la perturbación ambiental o el componente de estímulo externo, incluyendo modelos animales *in vivo* y/o experimentos de cultivo tisular *in vitro*.

10 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es el diagnóstico o la estadificación de un estado de enfermedad, en la que los moduladores identificados de un sistema biológico, por ejemplo, controladores de proceso celular determinantes (por ejemplo, diferenciales de interferencia o relaciones causales únicas en un sistema o proceso biológico) representan o bien marcadores de enfermedad o bien dianas terapéuticas que pueden estar sujetos a intervención terapéutica. La evaluación biológica por interrogación del sujeto es adecuada para cualquier estado de enfermedad en teoría, pero puede encontrarse que es particularmente útil en áreas tales como oncología/biología del cáncer, diabetes, obesidad, enfermedad cardiovascular y estados neurológicos (especialmente enfermedades neurodegenerativas, tales como, sin limitación, enfermedad de Alzheimer, enfermedad de Parkinson, enfermedad de Huntington, esclerosis lateral amiotrófica (ELA) y neurodegeneración relacionada con el envejecimiento).

20 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la determinación de la eficacia de un fármaco, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia o relaciones causales únicas en un sistema o proceso biológico) pueden ser los sellos distintivos de un fármaco exitoso y, a su vez, pueden usarse para identificar agentes adicionales, tales como MIM o intercambiadores epimetabólicos (*epishifters*), para tratar el mismo estado de enfermedad.

25 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la identificación de dianas farmacológicas para prevenir o tratar la infección, en la que el controlador de proceso celular determinante identificado (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores/indicadores o moléculas biológicas clave causantes del estado infeccioso y, a su vez, pueden usarse para identificar agentes antiinfecciosos.

30 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la evaluación de un efecto molecular de un agente, por ejemplo, un fármaco, sobre un perfil de enfermedad dado, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser un aumento o disminución de actividad de una o más rutas biológicas, o miembros clave de la(s) ruta(s), o reguladores clave para miembros de la(s) ruta(s), y que pueden usarse, a su vez, por ejemplo, para predecir la eficacia terapéutica del agente para la enfermedad dada.

40 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la evaluación del perfil toxicológico de un agente, por ejemplo, un fármaco, en una célula, un tejido, órgano u organismo, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser indicadores de toxicidad, por ejemplo, citotoxicidad y pueden usarse, a su vez, para predecir o identificar el perfil toxicológico del agente. En una realización, los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) es un indicador de cardiotoxicidad de un fármaco o candidato a fármaco y pueden usarse, a su vez, para predecir o identificar el perfil cardiotoxicológico del fármaco o candidato a fármaco.

50 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la identificación de dianas farmacológicas para prevenir o tratar una enfermedad o un trastorno provocado por armas biológicas, tales como protozoos, hongos, bacterias, protistas, virus o toxinas patógenos, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores/indicadores o moléculas biológicas clave causantes de dicha enfermedad o trastorno y pueden usarse, a su vez, para identificar agentes de biodefensa.

60 En determinados ejemplos que respaldan la comprensión, la evaluación biológica por interrogación es la identificación de dianas para agentes antienvjecimiento, tales como los cosméticos antienvjecimiento, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores o indicadores del proceso de envejecimiento, particularmente el proceso de envejecimiento en la piel y pueden usarse, a su vez, para identificar agentes antienvjecimiento.

65 En un modelo celular a modo de ejemplo para el envejecimiento que se usa en los métodos de la invención para identificar dianas para cosméticos antienvjecimiento, el modelo celular comprende una célula epitelial envejecida

que, por ejemplo, se trata con luz UV (una perturbación ambiental o un componente de estímulo externo), y/o células neonatales, que también se tratan opcionalmente con luz UV. En una realización, un modelo celular para el envejecimiento comprende un sistema celular de interferencia. En un sistema de interferencia de dos células a modo de ejemplo establecido para identificar dianas para cosméticos antienvjecimiento, una célula epitelial envejecida (primer sistema celular) puede tratarse con luz UV (un componente de estímulo externo) y pueden medirse cambios, por ejemplo, cambios proteómicos y/o cambios funcionales, en una célula neonatal (segundo sistema celular) resultante del contacto de las células neonatales con el medio condicionado de la célula epitelial envejecida tratada, por ejemplo, pueden medirse cambios proteómicos usando espectrometría de masas cuantitativa convencional, o puede identificarse una relación causal única en el envejecimiento a partir de una red de relaciones causales generada a partir de los datos.

En otro aspecto, la invención proporciona un kit para llevar a cabo una evaluación biológica por interrogación usando una tecnología de plataforma de descubrimiento, que comprende uno o más reactivos para detectar la presencia de y/o cuantificar la cantidad de, un analito que es el objeto de una causal red de relaciones generada a partir de los métodos de la invención. En una realización, dicho analito es el objeto de una relación causal única en el sistema biológico, por ejemplo, un gen asociado con una relación causal única en el sistema biológico. En determinadas realizaciones, el analito es una proteína, y los reactivos comprenden un anticuerpo contra la proteína, una etiqueta para la proteína y/o uno o más agentes para preparar la proteína para análisis de alto rendimiento (por ejemplo, secuenciación basada en espectrometría de masas).

Debe entenderse que todas las realizaciones descritas en el presente documento, incluyendo las descritas sólo en ejemplos, forman parte de la descripción general de la invención, y pueden combinarse con cualquier otra realización de la invención a menos que se renuncie a ello explícitamente o sea inaplicable.

Breve descripción de los dibujos

A continuación se describirán diversas realizaciones de la presente descripción con referencia a las figuras, en las que:

Figura 1: Ilustración del enfoque para identificar agentes terapéuticos.

Figura 2: Ilustración de la biología del cáncer del sistema y consecuencia de la regulación de salida interactiva multifisiológica integrada.

Figura 3: Ilustración de la interrogación sistemática de relevancia biológica usando MIMS.

Figura 4: Ilustración de modelado de red del cáncer para permitir la consulta biológica por interrogación.

Figura 5: Ilustración de la tecnología de plataforma de biología de interrogación.

Figura 6: Ilustración de tecnologías empleadas en la tecnología de plataforma.

Figura 7: Representación esquemática de los componentes de la plataforma, incluyendo la recopilación de datos, la integración de datos y la minería de datos.

Figura 8: Representación esquemática del interrogatorio sistemático usando MIMS y la recopilación de datos de respuesta de la cascada de "ómica".

Figura 9: Esquema de los componentes empleados para construir los modelos *in vitro* que representan estados normal y diabético.

Figura 10: Representación esquemática de la plataforma informática REFS™ usada para generar redes causales de la proteína en relación con la fisiopatología de la enfermedad.

Figura 11: Representación esquemática del enfoque para la generación de red diferencial en estados diabético frente a normal y ganglios de diabéticos que se restauran a estados normales mediante el tratamiento con MIMS.

Figura 12: Una red diferencial representativa en estados diabético frente a normal.

Figura 13: Una representación esquemática de un nodo y aristas de interés asociadas (nodo 1 en el centro). Se representa la funcionalidad celular asociada con cada arista.

Figura 14: Diagrama de flujo de alto nivel de un método a modo de ejemplo, según algunas realizaciones.

Figura 15A-15D: Ilustración esquemática de alto nivel de los componentes y procesos para un sistema de informática basado en IA que puede usarse con realizaciones a modo de ejemplo.

- Figura 16: Diagrama de flujo del proceso en un sistema de informática basado en IA que puede usarse con algunas realizaciones a modo de ejemplo.
- 5 Figura 17: Representa esquemáticamente un entorno informático a modo de ejemplo adecuado para poner en práctica realizaciones a modo de ejemplo enseñadas en el presente documento.
- Figura 18: Ilustración del diseño del estudio de casos descrito en el ejemplo 1.
- 10 Figura 19: Efecto de los tratamientos con CoQ10 sobre nodos posteriores.
- Figura 20: El tratamiento con CoQ10 disminuye la expresión de LDHA en la línea celular de cáncer HepG2.
- 15 Figura 21: Red de consenso de interacción proteica a modo de ejemplo a una frecuencia de fragmentos del 70% basándose en datos de líneas celulares Paca2, HepG2 y THLE2.
- Figura 22: Se identificaron proteínas que responden a la simulación de expresión de LDHA en dos líneas celulares de cáncer usando la tecnología de plataforma.
- 20 Figura 23: El análisis mediante Ingenuity Pathway Assist® de la red LDHA - PARK7 identifica a TP53 como centro anterior.
- Figura 24: Efecto del tratamiento con CoQ10 sobre los niveles de expresión de TP53 en la línea celular de cáncer SKMEL28.
- 25 Figura 25: La activación de TP53 asociada con la expresión alterada de las proteínas BCL-2 que efectúan apoptosis en la línea celular de cáncer SKMEL28 y efecto del tratamiento con CoQ10 sobre los niveles de expresión de Bcl-2, Bax y caspasa3 en SKMEL28.
- 30 Figura 26: Ilustración del enfoque matemático para la generación de redes delta-delta.
- Figura 27: Red (delta-delta) de diferencial cáncer-sano que controla TAEC y TCO. Cada controlador tiene efectos diferenciales sobre el punto final según lo representado por el grosor de la arista. El grosor de la arista en Cytoscape representa la intensidad del cambio en veces.
- 35 Figura 28: Mapeo de PARK7 y los nodos asociados de las salidas de tecnología de plataforma de interrogación usando IPA: Las formas grises incluyen todos los nodos asociados con PARK7 de las salidas de biología de interrogación que se importaron a IPA. Las formas sin rellenar (con nombres) son nuevas conexiones incorporadas por IPA para crear un mapa completo.
- 40 Figura 29: La tecnología de plataforma por interrogación de la invención, que demuestra asociaciones nuevas de nodos asociados con PARK7. Las aristas que se muestran en líneas discontinuas son conexiones entre dos nodos en las simulaciones que tienen nodos intermedios, pero no tienen nodos intermedios en IPA. Las aristas que se muestran en líneas de puntos son conexiones entre dos nodos en las simulaciones que tienen nodos intermedios, pero tienen nodos intermedios diferentes en IPA.
- 45 Figura 30: Ilustración del enfoque matemático para la generación de redes delta-delta. Compárense las aristas únicas de NG en la red delta $NG \cap HG$ con las aristas únicas de HGT1 en la red delta $HG \cap HGT1$. Las aristas en la intersección de NG y HGT1 son aristas HG que se restauran a NG con T1.
- 50 Figura 31: Red delta-delta de aristas de diabéticos restaurados a la normalidad con el tratamiento con coenzima Q10 superpuesto sobre la red delta $NG \cap HG$.
- Figura 32: Red delta-delta de aristas de pacientes hiperlipidémicos restaurados a la normalidad con el tratamiento con coenzima Q10 superpuesto sobre la red delta de hiperlipidemia \cap lipidemia normal.
- 55 Figura 33: Un esquema que representa el destino alterado del ácido graso en la enfermedad y el tratamiento farmacológico. Un equilibrio entre la utilización de ácidos grasos libres (AGL) para la generación de ATP y la remodelación de la membrana en respuesta a la perturbación de la biología de membrana se ha implicado en la cardiotoxicidad inducida por fármacos.
- 60 Figura 34: Un esquema que representa el diseño experimental y los parámetros de modelado usados para estudiar la toxicidad inducida por fármacos en cardiomiocitos de diabéticos.
- 65 Figura 35: La desregulación de la red transcripcional y la expresión de genes del metabolismo energético mitocondrial humano en cardiomiocitos de diabéticos por tratamiento farmacológico (T): la molécula de rescate (R)

normaliza la expresión génica.

Figura 36: A. El tratamiento farmacológico (T) indujo la expresión de GPAT1 y TAZ en mitocondrias de cardiomiocitos acondicionados en hiperglucemia. En combinación con la molécula de rescate (T+R), los niveles de GPAT1 y TAZ se normalizaron. B. Síntesis de TAG a partir de G3P.

Figura 37: A. El tratamiento farmacológico (T) disminuye la TCO (tasa de consumo de oxígeno) mitocondrial en cardiomiocitos acondicionados en hiperglucemia. La molécula de rescate (T+R) normaliza la TCO. B. El tratamiento farmacológico (T) reprime la síntesis de ATP mitocondrial en cardiomiocitos acondicionados en hiperglucemia.

Figura 38: Anotación de GO de proteínas reguladas por disminución mediante tratamiento farmacológico. Las proteínas implicadas en el metabolismo energético mitocondrial se regularon por disminución con el tratamiento farmacológico.

Figura 39: Ilustración del enfoque matemático para la generación de redes delta. Compárense las aristas únicas de T frente a UT, estando ambos modelos en un entorno diabético.

Figura 40: Un esquema que representa potenciales centros y redes de proteína que controlan la fisiopatología de la toxicidad inducida por fármacos.

Descripción detallada de la invención

I. Descripción general

Las realizaciones a modo de ejemplo de la presente invención incorporan métodos que pueden realizarse usando una plataforma de biología de interrogación ("la plataforma") que es una herramienta para comprender una amplia variedad de procesos biológicos, tales como la fisiopatología de la enfermedad, y los controladores moleculares clave subyacentes a tales procesos biológicos, incluyendo los factores que permiten un proceso de enfermedad. Algunas realizaciones a modo de ejemplo incluyen sistemas que pueden incorporar al menos una parte de, o la totalidad de, la plataforma. Algunos métodos a modo de ejemplo pueden emplear al menos parte de, o la totalidad de, la plataforma. Las metas y los objetivos de algunas realizaciones a modo de ejemplo que implican la plataforma se exponen en resumen de manera general a continuación con propósitos ilustrativos:

i) crear firmas moleculares específicas como controladores de componentes críticos del proceso biológico (por ejemplo, proceso de enfermedad) en relación con la fisiopatología general del proceso biológico;

ii) para generar firmas moleculares o mapas diferenciales pertenecientes al proceso biológico, que pueden ayudar a identificar firmas moleculares diferenciales que distinguen un estado biológico (por ejemplo, un estado de enfermedad) frente a un estado biológico diferente (por ejemplo, un estado normal), y desarrollar una comprensión de las firmas o entidades moleculares en la medida que arbitran mecanismos de cambio entre los dos estados biológicos (por ejemplo, del estado normal al de enfermedad); y,

iii) investigar el papel de los "centros" de actividad molecular como posibles dianas de intervención para el control externo del proceso biológico (por ejemplo, para usar el centro como un posible diana terapéutica), o como posibles marcadores biológicos para el proceso biológico en cuestión (por ejemplo, biomarcadores específicos de enfermedad, en usos de pronóstico y/o tratánósticos).

Algunos métodos a modo de ejemplo que implican a la plataforma pueden incluir una o más de las siguientes características:

1) modelar el proceso biológico (por ejemplo, proceso de enfermedad) y/o componentes del proceso biológico (por ejemplo, fisiología y fisiopatología de la enfermedad) en uno o más modelos, preferiblemente modelos *in vitro*, usando células asociadas con el proceso biológico. Por ejemplo, las células pueden ser células derivadas de humanos que participan normalmente en el proceso biológico en cuestión. El modelo puede incluir diversas señales/condiciones/perturbaciones celulares que son específicas del proceso biológico (por ejemplo, enfermedad). De manera ideal, el modelo representa diversos estados (enfermedad) y componentes de flujo, en lugar de una evaluación estática del estado biológico (enfermedad).

2) obtener perfiles de ARNm y/o firmas de proteína usando cualquier medio reconocido en la técnica. Por ejemplo, la reacción en cadena de la polimerasa cuantitativa (qPCR) y las herramientas de análisis de proteómica tales como espectrometría de masas (EM). Tales grupos de datos de ARNm y proteínas representan una reacción biológica al entorno/perturbación. Cuando corresponda y sea posible, los datos de lipidómica, metabolómica y transcriptómica también pueden integrarse como medidas complementarias o alternativas para el proceso biológico en cuestión. El análisis de SNP es otro componente que puede usarse a veces en el proceso. Puede ser útil para investigar, por ejemplo, si el SNP o una mutación específica tiene algún efecto sobre el proceso biológico. Estas variables pueden usarse para describir el

proceso biológico, o bien como una “instantánea” estática o bien como una representación de un proceso dinámico.

3) someter a ensayo una o más respuestas celulares a señales y perturbaciones, incluyendo pero sin limitarse a obtención de perfiles bioenergéticos, proliferación celular, apoptosis y función de orgánulo. La verdadera asociación genotipo-fenotipo se actualiza mediante el empleo de modelos funcionales, tales como los ensayos de ATP, ROS, OXPHOS, Seahorse, etc. Tales respuestas celulares representan la reacción de las células en el proceso biológico (o modelos del mismo) en respuesta al/a los estado(s) correspondiente(s) de la expresión de ARNm/proteína, y cualquier otro estado relacionado en el punto 2) anterior.

4) integrar datos de ensayos funcionales así obtenidos en el punto 3) con datos de proteómica y otros datos obtenidos en el punto 2), y determinar asociaciones proteicas controladas por causalidad, mediante el uso de una plataforma o un sistema de informática basado en inteligencia artificial (basado en IA). Tal sistema basado en IA se basa en, y preferiblemente se basa sólo en, los grupos de datos obtenidos en los puntos 2) y/o 3), sin recurrir al conocimiento existente sobre el proceso biológico. Preferiblemente, ningún punto de datos se corta estadística o artificialmente. En cambio, todos los datos obtenidos se alimentan al sistema de AI para determinar las asociaciones proteicas. Uno de los objetivos o salidas del proceso de integración es una o más redes diferenciales (de lo contrario, pueden denominarse en el presente documento “redes delta” o, en algunos casos, “redes delta-delta” dado el caso) entre los diferentes estados biológicos (por ejemplo, estados de enfermedad frente a normal).

5) obtener perfiles de las salidas de la plataforma de informática basada en IA para explorar cada centro de actividad como una posible diana terapéutica y/o biomarcador. Tal obtención de perfiles puede realizarse completamente *in silico* basándose en los grupos de datos obtenidos, sin recurrir a ningún experimento real de laboratorio de investigación.

6) validar el centro de actividad mediante el empleo de técnicas moleculares y celulares. Tal validación posinformática de la salida con experimentos basados en células de laboratorio de investigación puede ser opcional, pero ayudan a crear un círculo completo de interrogación.

Cualquiera o todos los enfoques expuestos en resumen antes pueden usarse en cualquier aplicación específica referente a cualquier proceso biológico dependiendo, al menos en parte, de la naturaleza de la aplicación específica. Es decir, uno o más enfoques expuestos en resumen antes pueden omitirse o modificarse, y pueden emplearse uno o más enfoques adicionales, dependiendo de la aplicación específica.

Se proporcionan diversos esquemas que ilustran la plataforma. En particular, en la figura 1 se representa una ilustración de un enfoque a modo de ejemplo para identificar agentes terapéuticos usando la plataforma. En la figura 2 se representa una ilustración de la biología del cáncer de los sistemas y la consecuencia de la regulación de salida interactiva multifisiológica integrada. En la figura 3 se representa una ilustración de interrogatorio sistemático de relevancia biológica usando MIMS. En la figura 4 se representa una ilustración de modelado de una red de cáncer para permitir una consulta biológica por interrogación. En las figuras 5 y 6 se representan ilustraciones de la plataforma de biología de interrogación y las tecnologías empleadas en la plataforma. En la figura 7 se presenta una representación esquemática de los componentes de la plataforma, incluyendo la recopilación de datos, la integración de datos y la minería de datos. En la figura 8 se presenta una representación esquemática de una interrogación sistemática usando MIMS y la recopilación de datos de respuesta de la cascada de “ómica”.)

La figura 14 es un diagrama de flujo de alto nivel de un método 10 a modo de ejemplo, en el que se indican los componentes de un sistema a modo de ejemplo que pueden usarse para realizar el método a modo de ejemplo. Inicialmente, se establece un modelo (por ejemplo, un modelo *in vitro*) para un proceso biológico (por ejemplo, un proceso de enfermedad) y/o componentes del proceso biológico (por ejemplo, fisiología y fisiopatología de la enfermedad) usando células asociadas normalmente con el proceso biológico (etapa 12). Por ejemplo, las células pueden ser células derivadas de humanos que participan normalmente en el proceso biológico (por ejemplo, enfermedad). El modelo celular puede incluir diversas señales celulares, condiciones y/o perturbaciones que son específicos del proceso biológico (por ejemplo, enfermedad). De manera ideal, el modelo celular representa diversos estados (enfermedad) y componentes de flujo del proceso biológico (por ejemplo, enfermedad), en lugar de una evaluación estática del proceso biológico. El modelo celular de comparación puede incluir células de control o células normales (por ejemplo, no enfermas). Una descripción adicional de los modelos celulares aparece a continuación en las secciones III.A y IV.

Se obtiene un primer grupo de datos del modelo celular para el proceso biológico, que incluye información que representa los niveles de expresión de una pluralidad de genes (por ejemplo, ARNm y/o firmas de proteína) (etapa 16) usando cualquier proceso o sistema conocido (por ejemplo, reacción en cadena de la polimerasa cuantitativa (qPCR) y herramientas de análisis de proteómica tales como espectrometría de masas (EM).

Se obtiene un tercer grupo de datos del modelo celular de comparación para el proceso biológico (etapa 18). El

tercer grupo de datos incluye información que representa los niveles de expresión de una pluralidad de genes en las células de comparación del modelo celular de comparación.

5 En determinadas realizaciones de los métodos de la invención, estos grupos de datos primero y tercero se denominan colectivamente en el presente documento un "primer grupo de datos" que representa los niveles de expresión de una pluralidad de genes en las células (todas las células, incluyendo las células de comparación) asociadas con el sistema biológico.

10 El primer grupo de datos y el tercer grupo de datos pueden obtenerse de uno o más Sistema(s) de análisis de firma de proteína y/o ARNm. Los datos de ARNm y proteína en los grupo de datos primero y tercero pueden representar reacciones biológicas al entorno y/o la perturbación. Cuando sea aplicable y posible, también pueden integrarse datos de lipidómica, metabolómica y transcriptómica como medidas complementarias o alternativas para el proceso biológico. El análisis de SNP es otro componente que puede usarse a veces en el proceso. Puede ser útil para investigar, por ejemplo, si un polimorfismo de un solo nucleótido (SNP) o una mutación específica tiene algún efecto sobre el proceso biológico. Las variables de datos pueden usarse para describir el proceso biológico, o bien como una "instantánea" estática o bien como una representación de un proceso dinámico. Una descripción adicional referente a la obtención de información que representa los niveles de expresión de una pluralidad de genes en las células aparece a continuación en la sección III.B.

20 Se obtiene un segundo grupo de datos del modelo celular para el proceso biológico, que incluye información que representa una actividad funcional o respuesta de las células (etapa 20). De manera similar, se obtiene un cuarto grupo de datos del modelo celular de comparación para el proceso biológico, que incluye información que representa una actividad funcional o respuesta de las células de comparación (etapa 22).

25 En determinadas realizaciones de los métodos de la invención, estos grupos de datos segundo y cuarto se denominan colectivamente en el presente documento un "segundo grupo de datos" que representa una actividad funcional o una respuesta celular de las células (todas las células, incluyendo las células de comparación) asociadas con el sistema biológico.

30 Pueden usarse uno o más sistemas de ensayo funcional para obtener información referente a la actividad funcional o la respuesta de las células o de las células de comparación. La información referente a las respuestas celulares funcionales a las señales y perturbaciones puede incluir, pero no se limita a, la obtención de perfiles bioenergéticos, la proliferación celular, la apoptosis y la función de orgánulo. Pueden emplearse modelos funcionales para procesos y rutas (por ejemplo, adenosina trifosfato (ATP), especies reactivas de oxígeno (ROS), fosforilación oxidativa (OXPHOS), ensayos de Seahorse, etc.) para obtener una verdadera asociación genotipo-fenotipo. La actividad funcional o las respuestas celulares representan la reacción de las células en el proceso biológico (o modelos del mismo) en respuesta al/a los estado(s) correspondiente(s) de la expresión de ARNm/proteína, y cualquier otra condición o perturbación aplicada relacionada. Se proporciona información adicional sobre la obtención de información que representa la actividad funcional o la respuesta de las células a continuación en la sección III.B.

40 El método también incluye generar modelos implementados por ordenador de los procesos biológicos en las células y en las células de control. Por ejemplo, pueden generarse una o más (por ejemplo, un conjunto de) redes bayesianas de relaciones causales entre el nivel de expresión de la pluralidad de genes y la actividad funcional o la respuesta celular para el modelo celular (las "redes de modelo celular generadas") a partir del primer grupo de datos y el segundo grupo de datos (etapa 24). Las redes de modelo celular generadas, individual o colectivamente, incluyen información direccional probabilística cuantitativa referente a las relaciones. Las redes de modelo celular generadas no se basan en relaciones biológicas conocidas entre la expresión génica y/o la actividad funcional o la respuesta celular, distinta de la información del primer grupo de datos y el segundo grupo de datos. La una o más redes de modelo celular generadas pueden denominarse colectivamente una red de modelo celular de consenso.

50 Pueden generarse una o más (por ejemplo, un conjunto de) redes bayesianas de relaciones causales entre el nivel de expresión de la pluralidad de genes y la actividad funcional o la respuesta celular para el modelo celular de comparación (las "redes de modelo celular de comparación generadas") a partir del primer grupo de datos y el segundo grupo de datos (etapa 26). Las redes de modelo celular de comparación generadas, individual o colectivamente, incluyen información direccional probabilística cuantitativa referente a las relaciones. Las redes celulares generadas no se basan en relaciones biológicas conocidas entre la expresión génica y/o la actividad funcional o la respuesta celular, distintas de la información en el primer grupo de datos y el segundo grupo de datos. La una o más redes de modelo de comparación generadas pueden denominarse colectivamente una red de modelo celular de consenso.

60 Las redes de modelo celular generadas y las redes de modelo celular de comparación generadas pueden crearse usando una plataforma de informática basada en inteligencia artificial (basada en IA). Más detalles referentes a la creación de las redes de modelo celular generadas, la creación de las redes de modelo celular de comparación generadas y el sistema de informática basado en IA aparecen a continuación en la sección III.C y en la descripción de las figuras 2A-3.

65

- 5 Debe observarse que pueden emplearse muchas plataformas o sistemas basados en IA diferentes para generar las redes bayesianas de relaciones causales, incluyendo información direccional probabilística cuantitativa. Aunque determinados ejemplos descritos en el presente documento emplean un sistema específico disponible comercialmente, es decir, REFS™ (ingeniería inversa/simulación directa) de GNS (Cambridge, MA), las realizaciones no están limitadas. Las plataformas o los sistemas basados en IA adecuados para implementar algunas realizaciones emplean algoritmos matemáticos para establecer relaciones causales entre las variables de entrada (por ejemplo, los grupo de datos primero y segundo), basándose sólo en los datos de entrada sin tener en cuenta el conocimiento previo existente sobre cualquier relación biológica potencial, establecida y/o verificada.
- 10 Por ejemplo, la plataforma de informática basada en AI REFS™ utiliza datos biológicos de entrada sin procesar (originales) o procesados mínimamente derivados experimentalmente (por ejemplo, datos genéticos, genómicos, epigenéticos, proteómicos, metabólicos y clínicos), y realiza rápidamente trillones de cálculos para determinar cómo interactúan las moléculas entre sí en un sistema completo. La plataforma de informática basada en AI REFS™ realiza un proceso de ingeniería inversa dirigido a crear un modelo celular implementado por ordenador *in silico* (por ejemplo, redes de modelo celular generadas), basándose en los datos de entrada, que representa cuantitativamente el sistema biológico subyacente. Además, pueden desarrollarse hipótesis sobre el sistema biológico subyacente y simularse rápidamente basándose en el modelo celular implementado por ordenador, para obtener predicciones, acompañadas de niveles de confianza asociados, con respecto a las hipótesis.
- 15
- 20 Con este enfoque, los sistemas biológicos están representados por modelos celulares cuantitativos implementados por ordenador en los que se simulan “intervenciones” para aprender mecanismos detallados del sistema biológico (por ejemplo, enfermedad), estrategias de intervención efectivas y/o biomarcadores clínicos que determinan qué pacientes responderán a un régimen de tratamiento dado. Los enfoques bioinformáticos y estadísticos convencionales, así como los enfoques basados en el modelado de la biología conocida, no son capaces normalmente de proporcionar este tipo de conocimientos.
- 25
- Después de crear las redes de modelo celular generadas y las redes de modelo celular de comparación generadas, se comparan. Se identifican una o más relaciones causales presentes en al menos algunas de las redes de modelo celular generadas, y están ausentes o tienen al menos un parámetro significativamente diferente en las redes de modelo celular de comparación generadas (etapa 28). Tal comparación puede dar como resultado la creación de una red diferencial. La comparación, identificación y/o creación de red diferencial (delta) pueden llevarse a cabo usando un módulo de creación de red diferencial, que se describe con más detalle a continuación en la sección III.D y con respecto a la descripción de la figura 26.
- 30
- 35 En algunas realizaciones, los grupos de datos de entrada son de un tipo de célula y un tipo de célula de comparación, lo que crea un conjunto de redes de modelo celular basándose en un tipo de célula y otro conjunto de redes de modelo celular de comparación basándose en un tipo de célula de control de comparación. Puede realizarse un diferencial entre el conjunto de redes del primer tipo de célula y el conjunto de redes del tipo o tipos de células de comparación.
- 40
- 45 En otras realizaciones, los grupos de datos de entrada son de múltiples tipos de células (por ejemplo, dos o más tipos de células cancerosas) y múltiples tipos de células de comparación (por ejemplo, dos o más tipos de células normales, no cancerosas). Puede generarse un conjunto de redes de modelo celular para cada tipo de célula y cada tipo de célula de comparación individualmente, y/o los datos de los múltiples tipos de célula y los múltiples tipos de célula de comparación pueden combinarse en grupos de datos compuestos respectivos. Los grupos de datos compuestos producen un conjunto de redes que corresponden a los múltiples tipos de células (datos compuestos) y otro conjunto de redes que corresponden a los múltiples tipos de células de comparación (datos compuestos de comparación). Puede realizarse un diferencial en el conjunto de redes para los datos compuestos en comparación con el conjunto de redes para los datos compuestos de comparación.
- 50
- En algunas realizaciones, puede realizarse un diferencial entre dos redes diferenciales diferentes. Esta salida puede denominarse una red delta-delta, y se describe a continuación con respecto a la figura 26.
- 55
- Puede identificarse información de relación cuantitativa para cada relación en las redes de modelo celular generadas (etapa 30). De manera similar, puede identificarse información de relación cuantitativa para cada relación en las redes de modelo celular de comparación generadas (etapa 32). La información cuantitativa referente a la relación puede incluir una dirección que indica causalidad, una medida de la incertidumbre estadística con respecto a la relación (por ejemplo, una medición estadística de área bajo la curva (AUC)), y/o una expresión de la magnitud cuantitativa de la intensidad de la relación (por ejemplo, una vez). Pueden obtenerse perfiles de las diversas relaciones en las redes de modelo celular generadas usando la información de relación cuantitativa para explorar cada centro de actividad en las redes como una posible diana terapéutica y/o biomarcador. Tal obtención de perfiles puede realizarse completamente *in silico* basándose en los resultados de las redes de modelo celular generadas, sin recurrir a ningún experimento real de laboratorio de investigación.
- 60
- 65 En algunas realizaciones, puede validarse un centro de actividad en las redes empleando técnicas moleculares y celulares. No es necesario realizar tal validación posinformática de la salida con experimentos basados en células de

laboratorio de investigación, pero puede ayudar a crear un círculo completo de interrogación. La figura 15 presenta esquemáticamente una representación de alto nivel simplificada de la funcionalidad de un sistema de informática basado en IA a modo de ejemplo (por ejemplo, sistema de informática basado en AI REFS™) e interacciones entre el sistema basado en IA y otros elementos o partes de una plataforma de biología de interrogación ("la plataforma").

5 En la figura 15A, diversos grupos de datos obtenidos de un modelo para un proceso biológico (por ejemplo, un modelo de enfermedad), tales como dosificación de fármaco, dosificación de tratamiento, expresión de proteína, expresión de ARNm y cualquiera de las muchas medidas funcionales asociadas (tales como TCO, TAEC) se introducen en un sistema basado en IA. Tal como se muestra en la figura 15B, a partir de los grupos de datos de entrada, el sistema de IA crea una biblioteca de "fragmentos de red" que incluye variables (proteínas, lípidos y metabolitos) que controlan los mecanismos moleculares en el proceso biológico (por ejemplo, enfermedad), en un proceso denominado enumeración de fragmentos bayesianos (figura 15B).

En la figura 15C, el sistema basado en AI selecciona un subconjunto de los fragmentos de red en la biblioteca y construye una red de ensayo inicial a partir de los fragmentos. El sistema basado en AI también selecciona un subconjunto diferente de los fragmentos de red en la biblioteca para construir otra red de ensayo inicial. Eventualmente, se crea un conjunto de redes de ensayo iniciales (por ejemplo, 1000 redes) a partir de diferentes subconjuntos de fragmentos de red en la biblioteca. Este proceso puede denominarse muestreo de conjuntos paralelos. Cada red de ensayo en el conjunto evoluciona o se optimiza añadiendo, restando y/o sustituyendo fragmentos de redes adicionales de la biblioteca. Si se obtienen datos adicionales, los datos adicionales pueden incorporarse a los fragmentos de red en la biblioteca y pueden incorporarse al conjunto de redes de ensayo a través de la evolución de cada red de ensayo. Después de completarse el proceso de optimización/evolución, el conjunto de redes de ensayo puede describirse como redes de modelo celular generadas.

Tal como se muestra en la figura 15D, el conjunto de redes de modelo celular generadas puede usarse para simular el comportamiento del sistema biológico. La simulación puede usarse para predecir el comportamiento del sistema biológico ante los cambios en las condiciones, lo que puede verificarse experimentalmente usando experimentos basados en células o basados en animales, de laboratorio de investigación. Además, pueden extraerse parámetros cuantitativos de las relaciones en las redes de modelo celular generadas usando la funcionalidad de simulación aplicando perturbaciones simuladas a cada nodo individualmente mientras se observan los efectos en los otros nodos en las redes del modelo celular generado. Se proporcionan más detalles a continuación en la sección III.C.

El proceso automatizado de ingeniería inversa del sistema de informática basado en IA, que se representa en las figuras 2A-2D, crea un conjunto de redes de modelo celular generadas que es un modelo de las células sin sesgo y sistemático basado en ordenador.

La ingeniería inversa determina las conexiones de red direccionales probabilísticas entre las mediciones moleculares en los datos y las salidas fenotípicas de interés. La variación en las mediciones moleculares permite aprender las relaciones probabilísticas de causa y efecto entre estas entidades y los cambios en los puntos finales. La naturaleza de aprendizaje automático de la plataforma también permite el entrenamiento cruzado y las predicciones basándose en un grupo de datos que está en constante evolución.

Las conexiones de red entre las mediciones moleculares en los datos son "probabilísticas", en parte porque la conexión puede basarse en correlaciones entre los grupos de datos observados "aprendidos" por el algoritmo informático. Por ejemplo, si el nivel de expresión de la proteína X y el de la proteína Y están correlacionados positiva o negativamente, basándose en el análisis estadístico del grupo de datos, puede asignarse una relación causal para establecer una conexión de red entre las proteínas X e Y. La fiabilidad de tal supuesta relación causal puede definirse adicionalmente mediante una probabilidad de la conexión, que puede medirse mediante el valor de p (por ejemplo, $p < 0,1, 0,05, 0,01$, etc.).

Las conexiones de red entre las mediciones moleculares en los datos son "direccionales", en parte porque las conexiones de red entre las mediciones moleculares, según se determina mediante el proceso de ingeniería inversa, refleja la causa y el efecto de la relación entre el gen/proteína conectados, de tal manera que elevar el nivel de expresión de una proteína puede provocar que el nivel de expresión de la otra aumente o disminuya, dependiendo de si la conexión es estimuladora o inhibitoria.

Las conexiones de red entre las mediciones moleculares en los datos son "cuantitativas", en parte porque las conexiones de red entre las mediciones moleculares, según se determina mediante el proceso, pueden simularse *in silico*, basándose en el grupo de datos existente y las medidas probabilísticas asociadas al mismo. Por ejemplo, en las conexiones de red establecidas entre las mediciones moleculares, puede ser posible aumentar o disminuir teóricamente (por ejemplo, en 1, 2, 3, 5, 10, 20, 30, 50, 100 veces o más) el nivel de expresión de una proteína dada (o un "nodo" en la red) y simular cuantitativamente sus efectos sobre otras proteínas conectadas en la red.

Las conexiones de red entre las mediciones moleculares en los datos son "sin sesgo", al menos en parte porque no hay puntos de datos cortados estadística o artificialmente, y en parte porque las conexiones de red se basan sólo en datos de entrada, sin hacer referencia al conocimiento preexistente sobre el proceso biológico en cuestión.

Las conexiones de red entre las mediciones moleculares en los datos son “sistémicas” y (sin sesgo), en parte porque todas las conexiones potenciales entre todas las variables de entrada se han explorado de manera sistémica, por ejemplo, por parejas. La dependencia de la potencia informática para ejecutar este sondeo sistémico aumenta exponencialmente a medida que aumenta el número de variables de entrada.

En general, un conjunto de ~1.000 redes suele ser suficiente para predecir relaciones cuantitativas causales probabilísticas entre todas las entidades medidas. El conjunto de redes captura la incertidumbre en los datos y permite el cálculo de métricas de confianza para cada predicción del modelo. Las predicciones se generaron usando el conjunto de redes conjuntamente, en el que las diferencias en las predicciones de redes individuales en el conjunto representan el grado de incertidumbre en la predicción. Esta característica permite la asignación de métricas de confianza para las predicciones de la respuesta clínica generada a partir del modelo.

Una vez que los modelos se diseñan por ingeniería inversa, pueden realizarse más consultas de simulación en el conjunto de modelos para determinar los controladores moleculares clave para el proceso biológico en cuestión, tal como un estado de enfermedad.

En la figura 9 se representa un esquema de componentes empleados para construir modelos *in vitro* a modo de ejemplo que representan el estado normal y diabético. En la figura 10 se presenta una representación esquemática de una plataforma de informática a modo de ejemplo REFS™ usada para generar redes causales de la proteína en relación con la fisiopatología de la enfermedad. En la figura 11 se presenta una representación esquemática del enfoque a modo de ejemplo para la generación de diferencial la red en estados diabético frente a normal y los nodos de diabéticos que se restauran a estados normales mediante tratamiento con MIMS. En la figura 12 se representa una red diferencial representativa en estados diabético frente a normal. En la figura 13 se presenta una representación esquemática de un nodo y las aristas de interés asociados (Nodo1 en el centro) y la funcionalidad celular asociada con cada arista.

Habiéndose descrito la invención en general anteriormente, las secciones a continuación proporcionan una descripción más detallada de diversos aspectos o elementos de la invención general, junto con uno o más sistemas biológicos específicos que pueden analizarse usando los métodos del presente documento. Sin embargo, debe observarse que los sistemas biológicos específicos usados con propósitos de ilustración a continuación no son limitativos. Por el contrario, se pretende que otros sistemas biológicos distintos, incluyendo cualquier alternativa, modificación y equivalente de los mismos, puedan analizarse de manera similar usando la tecnología de plataforma objeto.

II Definiciones

Tal como se usa en el presente documento, se definen en el presente documento determinados términos destinados a definirse específicamente, pero que no se definen ya en otras secciones de la memoria descriptiva.

Los artículos “un(o)” y “una” se usan en el presente documento para referirse a uno o más de uno (es decir, al menos a uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa un elemento o más de un elemento.

El término “incluyendo” se usa en el presente documento para significar, y se usa indistintamente con la expresión “incluyendo pero sin limitarse a”.

El término “o” se usa en el presente documento para significar, y se usa indistintamente con el término “y/o”, a menos que el contexto indique claramente otra cosa.

El término “tal como” se usa en el presente documento para significar, y se usa indistintamente, con la expresión “tal como pero no limitado a”.

“Ruta metabólica” se refiere a una secuencia de reacciones mediadas por enzimas que transforman un compuesto en otro y proporcionan productos intermedios y energía para funciones celulares. La ruta metabólica puede ser lineal o cíclica o ramificada.

“Estado metabólico” se refiere al contenido molecular de un entorno celular, multicelular o tisular particular en un punto en el tiempo dado, medido mediante diversos indicadores químicos y biológicos en relación con un estado de salud o enfermedad.

El término “microalineamiento” se refiere a un alineamiento de polinucleótidos, oligonucleótidos, polipéptidos (por ejemplo, anticuerpos) o péptidos sintetizados sobre un sustrato, tal como papel, nailon u otro tipo de membrana, filtro, chip, portaobjetos de vidrio, o cualquier otro soporte sólido adecuado.

Los términos “trastornos” y “enfermedades” se usan de manera inclusiva y se refieren a cualquier desviación de la estructura o función normal de cualquier parte, órgano o sistema del cuerpo (o cualquier combinación de los

- 5 mismos). Una enfermedad específica se manifiesta por síntomas y signos característicos, incluyendo cambios biológicos, químicos y físicos, y a menudo se asocia con una variedad de otros factores incluyendo, pero sin limitarse a, factores demográficos, ambientales, laborales, genéticos e históricos a nivel médico. Determinados signos característicos, síntomas y factores relacionados pueden cuantificarse a través de una variedad de métodos para proporcionar información de diagnóstico importante.
- 10 El término “expresión” incluye el proceso mediante el cual se produce un polipéptido a partir de polinucleótidos, tales como ADN. El proceso puede implicar la transcripción de un gen en ARNm y la traducción de este ARNm en un polipéptido. Dependiendo del contexto en el que se usa, “expresión” puede referirse a la producción de ARN, proteína o ambos.
- 15 Los términos “nivel de expresión de un gen” o “nivel de expresión génica” se refieren al nivel de ARNm, así como al/ a los transcrito(s) naciente(s) de pre-ARNm, productos intermedios de procesamiento de transcrito, ARN(s) maduro(s) y productos de degradación, o el nivel de proteína, codificada por el gen en la célula.
- 20 El término “modulación” se refiere a la regulación por incremento (es decir, activación o estimulación), regulación por disminución (es decir, inhibición o supresión) de una respuesta, o las dos en combinación o separadas. Un “modulador” es un compuesto o molécula que modula y puede ser, por ejemplo, un agonista, antagonista, activador, estimulador, supresor o inhibidor.
- 25 El término “trolamina”, tal como se usa en el presente documento, se refiere a trolamina NF, trietanolamina, TEALAN®, TEAlan 99%, trietanolamina, al 99%, trietanolamina, NF o trietanolamina, al 99%, NF. Estos términos pueden usarse indistintamente en el presente documento.
- 30 El término “genoma” se refiere a la totalidad de la información genética de una entidad biológica (célula, tejido, órgano, sistema, organismo). Está codificado o bien en ADN o bien en ARN (en determinados virus, por ejemplo). El genoma incluye tanto los genes como las secuencias no codificantes del ADN.
- 35 El término “proteoma” se refiere al grupo completo de proteínas expresadas por un genoma, una célula, un tejido o un organismo en un momento dado. Más específicamente, puede referirse al grupo completo de proteínas expresadas en un tipo dado de células o un organismo en un momento dado en condiciones definidas. El proteoma puede incluir variantes de proteína debido, por ejemplo, al corte y empalme alternativo de genes y/o modificaciones postraduccionales (tales como glicosilación o fosforilación).
- 40 El término “transcriptoma” se refiere al grupo completo de moléculas de ARN transcritas, incluyendo ARNm, ARNr, ARNt y otro ARN no codificante producido en una o una población de células en un momento dado. El término puede aplicarse al grupo total de transcritos en un organismo dado, o al subconjunto específico de transcritos presentes en un tipo de célula particular. A diferencia del genoma, que se fija aproximadamente para una línea celular dada (excluyendo mutaciones), el transcriptoma puede variar según las condiciones ambientales externas. Debido a que incluye todos los transcritos de ARNm en la célula, el transcriptoma refleja los genes que se expresan activamente en cualquier momento dado, con la excepción de los fenómenos de degradación de ARNm tales como atenuación transcripcional.
- 45 El estudio de transcriptómica, también denominado obtención de perfiles de expresión, examina el nivel de expresión de los ARNm en una población celular dada, a menudo usando técnicas de alto rendimiento basadas en la tecnología de microalineamientos de ADN.
- 50 El término “metaboloma” se refiere al grupo completo de metabolitos de molécula pequeña (tales como productos intermedios metabólicos, hormonas y otras moléculas de señalización, y metabolitos secundarios) que van a hallarse dentro de una muestra biológica, tal como un organismo individual, en un momento dado en una condición dada. El metaboloma es dinámico y puede cambiar de un segundo a otro.
- 55 El término “lipidoma” se refiere al grupo completo de lípidos que van a hallarse dentro de una muestra biológica, como un organismo individual, en un momento dado en una condición dada. El lipidoma es dinámico y puede cambiar de un segundo a otro.
- 60 El término “interactoma” se refiere al grupo completo de interacciones moleculares en un sistema biológico en estudio (por ejemplo, células). Puede mostrarse como un grafo dirigido. Pueden producirse interacciones moleculares entre moléculas que pertenecen a diferentes familias bioquímicas (proteínas, ácidos nucleicos, lípidos, hidratos de carbono, etc.) y también dentro de una familia dada. Cuando se habla en términos de proteómica, interactoma se refiere a la red de interacción proteína-proteína (IPP), o red de interacción proteica (RIP). Otro tipo de interactoma estudiado ampliamente es el interactoma proteína-ADN (red formada por factores de transcripción (y proteínas reguladoras de ADN o cromatina) y sus genes diana.
- 65 El término “salida celular” incluye una colección de parámetros, preferiblemente parámetros medibles, relacionados con el estado celular, incluyendo (sin limitación): nivel de transcripción para uno o más genes (por ejemplo, medible

mediante RT-PCR, qPCR, microalineamiento, etc.), nivel de expresión para una o más proteínas (por ejemplo, medible mediante espectrometría de masas o inmunotransferencia de tipo Western), actividad absoluta (por ejemplo, medible como tasas de conversión de sustrato) o actividad relativa (por ejemplo, medible como un valor en % en comparación con la actividad máxima) de una o más enzimas o proteínas, nivel de uno o más metabolitos o productos intermedios, nivel de fosforilación oxidativa (por ejemplo, medible mediante la tasa de consumo de oxígeno u TCO), nivel de glicólisis (por ejemplo, medible mediante la tasa de acidificación extracelular o TAEC), extensión de la unión o interacción ligando-diana, actividad de moléculas secretadas extracelulares, etc. La salida celular puede incluir datos para un número predeterminado de genes o proteínas diana, etc., o puede incluir una evaluación global para todos los genes o proteínas detectables. Por ejemplo, puede usarse espectrometría de masas para identificar y/o cuantificar todas las proteínas detectables expresadas en una muestra o población celular dada, sin conocimiento previo de si alguna proteína específica puede expresarse en la muestra o población celular.

Tal como se usa en el presente documento, un “sistema celular” incluye una población de células homogéneas o heterogéneas. Las células dentro del sistema pueden estar creciendo *in vivo*, en el entorno natural o fisiológico, o puede estar creciendo *in vitro*, por ejemplo, en entornos controlados de cultivo tisular. Las células dentro del sistema pueden ser relativamente homogéneas (por ejemplo, homogéneas en no menos del 70%, el 80%, el 90%, el 95%, el 99%, el 99,5%, el 99,9%), o pueden contener dos o más tipos de células, tales como los tipos de células que se encuentran generalmente en estrecha proximidad *in vivo*, o tipos de células que pueden interaccionar entre sí *in vivo* mediante, por ejemplo, comunicación intercelular a larga distancia paracrina u otra. Las células dentro del sistema celular pueden derivar de líneas celulares establecidas, incluyendo líneas celulares de cáncer, líneas celulares inmortales o líneas celulares normales, o pueden ser células primarias o células recién aisladas de tejidos u órganos vivos.

Las células en el sistema celular están normalmente en contacto con un “entorno celular” que puede proporcionar nutrientes, gases (oxígeno o CO₂, etc.), productos químicos o estimulantes proteínicos/no proteínicos que pueden definir las condiciones que afectan al comportamiento celular. El entorno celular puede ser un medio químico con componentes químicos definidos y/o extractos tisulares o componentes séricos peor definidos, y puede incluir un pH, contenido de CO₂, presión y temperatura específicos en los cuales crecen las células. Alternativamente, el entorno celular puede ser el entorno natural o fisiológico que se encuentra *in vivo* para el sistema celular específico.

En determinadas realizaciones, un entorno celular comprende condiciones que simulan un aspecto de un sistema o proceso biológico, por ejemplo, simulan un estado de enfermedad, proceso o entorno. Tales condiciones de cultivo incluyen, por ejemplo, condiciones de hiperglucemia, hipoxia o ricas en láctico. Numerosas otras condiciones se describen en el presente documento.

En determinadas realizaciones, un entorno celular para un sistema celular específico también incluye determinadas características de la superficie celular del sistema celular, tales como los tipos de receptores o ligandos en la superficie celular y sus actividades respectivas, la estructura de las moléculas de hidratos de carbono o lípidos, la polaridad o fluidez de la membrana, estado de agrupamiento de determinadas proteínas de membrana, etc. Estas características de la superficie celular pueden afectar a la función de las células cercanas, tales como células que pertenecen a un sistema celular diferente. Sin embargo, en determinadas otras realizaciones, el entorno celular de un sistema celular no incluye características de la superficie celular del sistema celular.

El entorno celular puede modificarse para convertirse en un “entorno celular modificado”. Las alteraciones pueden incluir cambios (por ejemplo, aumentar o disminuir) en uno o más componentes que se encuentran en el entorno celular, incluyendo la adición de uno o más “componentes de estímulo externo” al entorno celular. La perturbación ambiental o el componente de estímulo externo puede ser endógeno con respecto al entorno celular (por ejemplo, el entorno celular contiene ciertos niveles del estimulante, y se añade más de lo mismo para aumentar su nivel), o puede ser exógeno con respecto al entorno celular (por ejemplo, el estimulante está ausente en gran medida del entorno celular antes de la alteración). El entorno celular puede alterarse adicionalmente por cambios secundarios resultantes de la adición del componente de estímulo externo, puesto que el componente de estímulo externo puede cambiar la salida celular del sistema celular, incluyendo las moléculas secretadas en el entorno celular por el sistema celular.

Tal como se usa en el presente documento, el “componente de estímulo externo”, también denominado en el presente documento “perturbación ambiental”, incluye cualquier estímulo físico y/o químico externo que pueda afectar a la función celular. Esto puede incluir moléculas orgánicas o inorgánicas grandes o pequeñas, productos químicos naturales o sintéticos, cambio de temperatura, cambio de pH, radiación, luz (UVA, UVB, etc.), microondas, ondas sónicas, corriente eléctrica, campos magnéticos modulados o no modulados, etc.

El término “molécula intracelular multidimensional (MIM)”, es una versión aislada o una versión producida de manera sintética de una molécula endógena que se produce de manera natural por el cuerpo y/o está presente en al menos una célula de un humano. Una MIM es capaz de entrar en una célula y la entrada en la célula incluye la entrada total o parcial en la célula siempre que la parte biológicamente activa de la molécula entre por completo en la célula. Las MIM son capaces de inducir un mecanismo de transducción de señales y/o expresión génica dentro de una célula. Las MIM son multidimensionales porque las moléculas tienen tanto un efecto terapéutico como uno portador, por

ejemplo, de administración de fármaco. Las MIM también son multidimensionales porque las moléculas actúan de una manera en un estado de enfermedad y de una manera diferente en un estado normal. Por ejemplo, en el caso de CoQ-10, la administración de CoQ-10 a una célula de melanoma en presencia de VEGF conduce a un nivel disminuido de Bcl2 que conduce, a su vez, a un potencial oncogénico disminuido para la célula de melanoma. Por el contrario, en un fibroblasto normal, la coadministración de CoQ-10 y VEGF no tiene ningún efecto sobre los niveles de Bcl2.

En una realización, una MIM también es un intercambiador epimetabólico (*epi-shifter*). En otra realización, una MIM no es un intercambiador epimetabólico. En otra realización, una MIM se caracteriza por una o más de las funciones anteriores. En otra realización, una MIM se caracteriza por dos o más de las funciones anteriores. En una realización adicional, una MIM se caracteriza por tres o más de las funciones anteriores. En aún otra realización, una MIM se caracteriza por todas las funciones anteriores. El experto apreciará que una MIM de la invención también pretende abarcar una mezcla de dos o más moléculas endógenas, en el que la mezcla se caracteriza por una o más de las funciones anteriores. Las moléculas endógenas en la mezcla están presentes en una razón tal que la mezcla funciona como MIM.

Las MIM pueden ser moléculas basadas en lípidos o no basadas en lípidos. Los ejemplos de las MIM incluyen, pero no se limitan a, CoQ10, acetil Co-A, palmitil Co-A, L-carnitina, aminoácidos tales como, por ejemplo, tirosina, fenilalanina y cisteína. En una realización, la MIM es una molécula pequeña. En una realización de la invención, la MIM no es CoQ10. Las MIM pueden identificarse de manera rutinaria por un experto en la técnica usando cualquiera de los ensayos descritos con detalle en el presente documento. Las MIM se describen con más detalle en el documento US 12/777.902 (US 2011-0110914).

Tal como se usa en el presente documento, un “intercambiador epimetabólico” (*epi-shifter*) es una molécula que modula el intercambio metabólico de un estado sano (o normal) a un estado de enfermedad y viceversa, manteniendo o restableciendo de ese modo la salud de la célula, el tejido, sistema y/o huésped en un humano. Los intercambiadores epimetabólicos son capaces de efectuar la normalización en un microentorno tisular. Por ejemplo, un intercambiador epimetabólico incluye cualquier molécula que sea capaz, cuando se añade o se reduce de una célula, de afectar al microentorno (por ejemplo, el estado metabólico) de una célula. El experto apreciará que un intercambiador epimetabólico de la invención también pretende abarcar una mezcla de dos o más moléculas, en el que la mezcla se caracteriza por una o más de las funciones anteriores. Las moléculas en la mezcla están presentes en una razón tal que la mezcla funciona como intercambiador epimetabólico. Los ejemplos de intercambiadores epimetabólicos incluyen, pero no se limitan a, CoQ-10; vitamina D3; componentes de MEC tales como fibronectina; inmunomoduladores, tales como TNFa o cualquiera de las interleucinas, por ejemplo, IL-5, IL-12, IL-23; factores angiogénicos; y factores apoptóticos.

En una realización, el intercambiador epimetabólico también es una MIM. En una realización, el intercambiador epimetabólico no es CoQ10. Los intercambiadores epimetabólicos pueden identificarse de manera rutinaria por un experto en la técnica usando cualquiera de los ensayos descritos con detalle en el presente documento. Los intercambiadores epimetabólicos se describen con más detalle en el documento US 12/777.902 (US 2011-0110914).

Otros términos que no se definen explícitamente en la presente solicitud tienen un significado tal como lo habría entendido un experto en la técnica.

III. Etapas y componentes a modo de ejemplo de la tecnología de plataforma

Sólo con propósitos ilustrativos, las siguientes etapas de la tecnología de plataforma objeto pueden describirse a continuación en el presente documento como una utilidad a modo de ejemplo para integrar datos obtenidos de un modelo de cáncer personalizado y para identificar nuevas proteínas/rutas que controlan la patogenia del cáncer. Los mapas relacionales resultantes de este análisis proporcionan dianas de tratamiento del cáncer, así como marcadores de diagnóstico/pronóstico asociados al cáncer. Sin embargo, la tecnología de plataforma objeto tiene aplicabilidad general para cualquier sistema o proceso biológico, y no se limita a ningún cáncer en particular u otros modelos de enfermedad específicos.

Además, aunque la descripción a continuación se presenta en algunas partes como etapas diferenciadas, es con propósitos ilustrativos y de simplicidad, y por tanto, en realidad, no implica un orden y/o una demarcación de etapas tan rígidos. Además, las etapas de la invención pueden realizarse por separado, y la invención proporcionada en el presente documento pretende abarcar cada una de las etapas individuales por separado, así como combinaciones de una o más etapas (por ejemplo, una cualquiera, dos, tres, cuatro, cinco, seis o las siete etapas) de la tecnología de plataforma objeto, que puede llevarse a cabo independientemente de las etapas restantes.

La invención también pretende incluir todos los aspectos de la tecnología de plataforma como componentes independientes y realizaciones de la invención. Por ejemplo, los grupos de datos generados pretenden ser realizaciones de la invención. Como ejemplos adicionales, las redes de relaciones causales generadas, las redes de relaciones causales de consenso generadas, y/o las redes de relaciones causales simuladas generadas, también pretenden ser realizaciones de la invención. Las relaciones causales identificadas como únicas en el sistema

biológico pretenden ser realizaciones de la invención. Además, los modelos personalizados para un sistema biológico particular también pretenden ser realizaciones de la invención. Por ejemplo, los modelos personalizados para un estado o proceso de enfermedad tales como, por ejemplo, modelos celulares para cáncer, obesidad/diabetes/enfermedad cardiovascular, o un modelo personalizado para la toxicidad (por ejemplo, cardiotoxicidad) de un fármaco, también pretenden ser realizaciones de la invención.

A. Construcción de modelos personalizados

La primera etapa en la tecnología de plataforma es el establecimiento de un modelo para un sistema o proceso biológico. Un ejemplo de un sistema o proceso biológico es el cáncer. Como cualquier otro proceso o sistema biológico complicado, el cáncer es un estado patológico complicado caracterizado por múltiples aspectos únicos. Por ejemplo, debido a su alta velocidad de crecimiento, muchas células cancerosas están adaptadas para crecer en condiciones de hipoxia, tienen glicólisis regulada por incremento y rutas metabólicas de fosforilación oxidativa reducidas. Como resultado, las células cancerosas pueden reaccionar de manera diferente a una perturbación ambiental, tal como tratamiento con un fármaco potencial, en comparación con la reacción de una célula normal en respuesta al mismo tratamiento. Por tanto, sería interesante descifrar las respuestas únicas del cáncer al tratamiento farmacológico en comparación con las respuestas de las células normales. Con este propósito, puede establecerse un modelo de cáncer personalizado para simular el entorno de una célula cancerosa, por ejemplo, dentro de un tumor *in vivo*, creando condiciones de cultivo celular que se aproximen estrechamente a las condiciones de una célula cancerosa en un tumor *in vivo*, o para imitar diversos aspectos del crecimiento del cáncer, aislando diferentes condiciones de crecimiento de las células cancerosas.

Uno de estos "entornos" de cáncer, o condición de estrés de crecimiento, es la hipoxia, una condición que se encuentra normalmente dentro de un tumor sólido. Puede inducirse hipoxia en células usando métodos reconocidos en la técnica. Por ejemplo, puede inducirse hipoxia colocando sistemas celulares en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que puede inundarse con una mezcla de gases industriales que contiene el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Pueden medirse los efectos después de un periodo predeterminado, por ejemplo, a las 24 horas después del tratamiento de hipoxia, con y sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 μM).

Asimismo, el tratamiento con ácido láctico de las células imita un entorno celular en el que la actividad de glicólisis es alta, como existe en el entorno tumoral *in vivo*. El estrés inducido por ácido láctico puede investigarse a una concentración de ácido láctico final de aproximadamente 12,5 mM en un tiempo predeterminado, por ejemplo, a las 24 horas, con o sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 μM).

La hiperglucemia es normalmente una condición que se encuentra en la diabetes; sin embargo, la hiperglucemia también imita en cierta medida un aspecto del crecimiento del cáncer porque muchas células cancerosas dependen de la glucosa como su principal fuente de energía. La exposición de las células objeto a una condición hiperglucémica típica puede incluir la adición del 10% de glucosa de grado de cultivo a medios adecuados, de modo que la concentración final de glucosa en los medios sea de aproximadamente 22 mM.

Las condiciones individuales que reflejan diferentes aspectos del crecimiento del cáncer pueden investigarse por separado en el modelo de cáncer personalizado y/o pueden combinarse entre sí. En una realización, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más condiciones que reflejan o simulan diferentes aspectos del crecimiento/estados del cáncer se investigan en el modelo de cáncer personalizado. En una realización, las condiciones individuales y, además, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más de las condiciones que reflejan o simulan diferentes aspectos del crecimiento/estados del cáncer se investigan en el modelo de cáncer personalizado. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 ó 10 y 50 condiciones diferentes.

A continuación se enumeran algunas combinaciones a modo de ejemplo de condiciones que pueden usarse para tratar células. Pueden formularse fácilmente otras combinaciones dependiendo de la evaluación biológica por interrogación específica que esté realizándose.

1. Sólo medios
2. CTL Coenzima Q10 50 μM (CoQ10)
3. CTL Coenzima Q10100 μM
4. Ácido láctico 12,5 mM
5. Ácido láctico 12,5 mM + CTL coenzima Q10 50 μM

- 5 6. Ácido láctico 12,5 mM + CTL coenzima Q10 100 μ M
7. Hipoxia
8. Hipoxia + CTL coenzima Q10 50 μ M
9. Hipoxia + CTL coenzima Q10 100 μ M
- 10 10. Hipoxia + ácido láctico 12,5 mM
11. Hipoxia + ácido láctico 12,5 mM + CTL coenzima Q10 50 μ M
- 15 12. Hipoxia + ácido láctico 12,5 mM + CTL coenzima Q10 100 μ M
13. Medios + glucosa 22 mM
14. CTL coenzima Q10 50 μ M + glucosa 22 mM
- 20 15. CTL coenzima Q10 100 μ M + glucosa 22 mM
16. Ácido láctico 12,5 mM + glucosa 22 mM
- 25 17. Ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 50 μ M
18. Ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 100 μ M
19. Hipoxia + glucosa 22 mM
- 30 20. Hipoxia + glucosa 22 mM + CTL coenzima Q10 50 μ M
21. Hipoxia + glucosa 22 mM + CTL coenzima Q10 100 μ M
- 35 22. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM
23. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 50 μ M
24. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 100 μ M
- 40 Como control, se cultivan una o más líneas celulares normales (por ejemplo, THLE2 y HDFa) en condiciones similares para identificar proteínas o rutas únicas del cáncer (véase a continuación). El control puede ser el modelo celular de comparación descrito anteriormente.
- 45 Pueden incluirse en el modelo de cáncer múltiples células cancerosas del mismo o diferente origen (por ejemplo, líneas de cáncer PaCa2, HepG2, PC3 y MCF7), en oposición a un único tipo de célula cancerosa. En determinadas situaciones, pueden realizarse experimentos de interferencia o SEC entre diferentes células cancerosas (por ejemplo, HepG2 y PaCa2) con diversos propósitos interrelacionados.
- 50 En algunas realizaciones que implican interferencia, los experimentos realizados en los modelos celulares están diseñados para determinar la modulación del estado celular o la función de un sistema o una población celular (por ejemplo, célula de hepatocarcinoma HepG2) por otro sistema o población celular (por ejemplo, cáncer de páncreas PaCa2) en condiciones de tratamiento definidas (por ejemplo, hiperglucemia, hipoxia (isquemia)). Según una configuración típica, un primer sistema/población celular se pone en contacto con componentes de estímulo externo, tales como una molécula candidata (por ejemplo, una molécula de fármaco pequeña, una proteína) o una condición candidata (por ejemplo, hipoxia, entorno con alto contenido de glucosa). En respuesta, el primer sistema/población celular cambia su transcriptoma, proteoma, metaboloma y/o interactoma, lo que conduce a cambios que pueden detectarse fácilmente tanto en el interior como en el exterior de la célula. Por ejemplo, los cambios en el transcriptoma pueden medirse mediante el nivel de transcripción de una pluralidad de ARNm diana; los cambios en el proteoma pueden medirse mediante el nivel de expresión de una pluralidad de proteínas diana; y los cambios en el metaboloma pueden medirse mediante el nivel de una pluralidad de metabolitos diana mediante ensayos diseñados específicamente para metabolitos dados. Alternativamente, los cambios a los que se hizo referencia anteriormente en el metaboloma y/o proteoma, al menos con respecto a determinados metabolitos o proteínas secretados, también pueden medirse mediante sus efectos sobre el segundo sistema/población celular, incluyendo la modulación del transcriptoma, proteoma, metaboloma e interactoma del segundo sistema/población celular. Por tanto, los experimentos pueden usarse para identificar los efectos de las molécula(s) de interés secretada(s) por el
- 65

primer sistema/población celular sobre un segundo sistema/población celular en diferentes condiciones de tratamiento. Los experimentos también pueden usarse para identificar cualquier proteína que se module como resultado de la señalización del primer sistema celular (en respuesta al tratamiento con el componente de estímulo externo) a otro sistema celular, por ejemplo, mediante examen diferencial de proteómica. La misma configuración experimental también puede adaptarse para una configuración inversa, de modo que también puedan evaluarse los efectos recíprocos entre los dos sistemas celulares. En general, para este tipo de experimento, la elección de las parejas de líneas celulares se basa en gran medida en factores tales como origen, el estado de enfermedad y la función celular.

Aunque están implicados normalmente sistemas de dos células en este tipo de configuración experimental, también pueden diseñarse experimentos similares para más de dos sistemas celulares, por ejemplo, inmovilizando cada sistema celular distinto sobre un soporte sólido independiente.

Una vez que se construye el modelo personalizado, pueden aplicarse una o más "perturbaciones" al sistema, tales como variación genética de un paciente a otro, o con/sin tratamiento con determinados fármacos o profármacos. Véase la figura 15D. Los efectos de tales perturbaciones en el sistema, incluyendo el efecto sobre las células cancerosas relacionadas con la enfermedad y las células de control normales relacionadas con la enfermedad, pueden medirse usando diversos medios reconocidos en la técnica o patentados, tal como se describe en la sección III.B a continuación.

En un experimento a modo de ejemplo, las líneas de cáncer PaCa2, HepG2, PC3 y MCF7, y las líneas celulares normales THLE2 y HDFa, se acondicionan en cada uno de condiciones de hiperglucemia, hipoxia y ricas en ácido láctico, así como en todas las combinaciones de dos o tres de estas condiciones, y además con o sin perturbación ambiental, específicamente el tratamiento con coenzima Q10.

El modelo celular personalizado puede establecerse y usarse en la totalidad de las etapas de la tecnología de plataforma de la invención para identificar en última instancia una relación causal única en el sistema biológico, llevando a cabo las etapas descritas en el presente documento. Sin embargo, el experto en la técnica entenderá que un modelo celular personalizado que se usa para generar una red de relaciones causales de consenso inicial de "primera generación" para un proceso biológico puede evolucionar o expandirse de manera continua a lo largo del tiempo, por ejemplo, mediante la introducción de líneas celulares de cáncer o normales adicional y/o estados de cáncer adicionales. Pueden recopilarse datos adicionales del modelo celular evolucionado, es decir, datos de la(s) parte(s) recién añadida(s) del modelo celular. Los nuevos datos recopilados de un modelo celular expandido o evolucionado, es decir, de la(s) parte(s) recién añadida(s) del modelo celular, pueden introducirse luego en los grupos de datos usados previamente para generar la red de relaciones causales de consenso de "primera generación" con el fin de generar una red de relaciones causales de consenso de "segunda generación" más robusta. Pueden identificarse nuevas relaciones causales únicas del sistema biológico a partir de la red de relaciones causales de consenso de "segunda generación". De esta manera, la evolución del modelo celular proporciona una evolución de las redes de relaciones causales de consenso, proporcionando así conocimientos nuevos y/o más fiables sobre los moduladores del sistema biológico.

Se describen con detalle ejemplos adicionales de modelos de células personalizadas en el presente documento.

B. Recopilación de datos

En general, pueden recopilarse dos tipos de datos de cualquier sistema de modelo personalizado. Un tipo de datos (por ejemplo, el primer grupo de datos, el tercer grupo de datos) se relaciona generalmente con el nivel de determinadas macromoléculas, como ADN, ARN, proteínas, lípidos, etc. Un grupo de datos a modo de ejemplo en esta categoría son los datos proteómicos (por ejemplo, datos cualitativos y cuantitativos referentes a la expresión de todas o sustancialmente todas las proteínas medibles de una muestra). El otro tipo de datos son generalmente datos funcionales (por ejemplo, el segundo grupo de datos, el cuarto grupo de datos) que reflejan los cambios fenotípicos resultantes de los cambios en el primer tipo de datos.

Con respecto al primer tipo de datos, en algunas realizaciones de ejemplo, se realizan reacción en cadena de la polimerasa cuantitativa (qPCR) y proteómica para perfilar los cambios en la expresión de proteína y ARNm celular mediante reacción en cadena de la polimerasa cuantitativa (qPCR) y proteómica. El ARN total puede aislarse usando un kit de aislamiento de ARN comercial. Tras la síntesis de ADNc, pueden emplearse alineamientos de qPCR disponibles comercialmente específicos (por ejemplo, los de SA Biosciences) para el área de enfermedad o procesos celulares tales como angiogénesis, apoptosis y diabetes, para perfilar un grupo predeterminado de genes siguiendo las instrucciones del fabricante. Por ejemplo, el sistema de amplificación cfx-384 de Biorad puede usarse para todos los experimentos de obtención de perfiles transcripcionales. Tras la recopilación de datos (Ct), el cambio en veces final con respecto al control puede determinarse usando el método δ Ct tal como se describe en el protocolo del fabricante. El análisis de muestras proteómicas puede realizarse tal como se describe en las secciones posteriores.

El método objeto puede emplear análisis proteómico cuantitativo de alto rendimiento a gran escala de cientos de

muestras de carácter similar, y proporciona los datos necesarios para identificar los diferenciales de salida celular.

Existen numerosas tecnologías reconocidas en la técnica adecuadas con este propósito. A continuación se describe en resumen una técnica a modo de ejemplo, el análisis iTRAQ en combinación con la espectrometría de masas.

El enfoque de proteómica cuantitativa se basa en el etiquetado de isótopos estables con el reactivo iTRAQ 8-plex y 2D-CL MALDI EM/EM para la identificación y cuantificación de péptidos. La cuantificación con esta técnica es relativa: a los péptidos y las proteínas se les asignan proporciones de abundancia en relación con una muestra de referencia. Las muestras de referencia comunes en múltiples experimentos de iTRAQ facilitan la comparación de muestras en múltiples experimentos de iTRAQ.

Por ejemplo, para implementar este esquema de análisis, pueden combinarse seis muestras primarias y dos muestras de reserva de control en una mezcla de iTRAQ 8 plex según las sugerencias del fabricante. Esta mezcla de ocho muestras puede fraccionarse por cromatografía de líquidos bidimensional; intercambio catiónico fuerte (SCX) en la primera dimensión, y HPLC de fase inversa en la segunda dimensión, pueden someterse a un análisis mediante espectrometría de masas.

En el presente documento se proporciona una breve descripción de procedimientos de laboratorio a modo de ejemplo que pueden emplearse.

Extracción de proteína: Las células pueden lisarse con tampón de lisis de urea 8 M con inhibidores de proteasa (inhibidor de proteasa de parada de Thermo Scientific sin EDTA) e incubarse en hielo durante 30 minutos con vórtex durante 5 segundos cada 10 minutos. La lisis puede completarse mediante ultrasonidos en un pulso de 5 segundos. Los lisados celulares pueden centrifugarse a 14000 x g durante 15 minutos (4°C) para eliminar los restos celulares. El ensayo de Bradford puede realizarse para determinar la concentración de proteína. Pueden reducirse 100 g de proteína de cada muestra (ditiotreitól (DTT) 10 mM, 55°C, 1 h), alquilarse (yodoacetamida 25 mM, temperatura ambiente, 30 minutos) y digerirse con tripsina (1:25 p/p, bicarbonato de trietilamonio (TEAB) 200 mM, 37°C, 16 h).

Preparación de la muestra de secretoma: 1) En una realización, las células pueden cultivarse en medio libre de suero: pueden concentrarse medios condicionados mediante liofilizador, reducirse (ditiotreitól (DTT) 10 mM, 55°C, 1 h), alquilarse (yodoacetamida 25 mM, a temperatura ambiente, incubarse durante 30 minutos), y luego desalarse mediante precipitación con actona. Puede digerirse la misma cantidad de proteínas de los medios condicionados concentrados, con tripsina (1:25 p/p, bicarbonato de trietilamonio (TEAB) 200 mM, 37°C, 16 h).

En una realización, las células pueden cultivarse en medio que contiene suero: el volumen del medio puede reducirse usando columnas de Vivaspín de MWCO 3k (GE Healthcare Life Sciences), luego puede reconstituirse con 1xPBS (Invitrogen). La albúmina sérica puede reducirse de todas las muestras usando la columna AlbuVoid (Biotech Support Group, LLC) siguiendo las instrucciones del fabricante con las modificaciones del intercambio de tampón para optimizar la aplicación del medio de condición.

Etiquetado con iTRAQ 8-plex: Puede agruparse una alícuota de cada digesto trípico en cada grupo experimental para crear la muestra de control agrupada. Pueden etiquetarse alícuotas iguales de cada muestra y la muestra de control agrupada con reactivos de iTRAQ 8-plex según los protocolos del fabricante (AB Sciex). Las reacciones pueden combinarse, aspirar a vacío hasta sequedad, resuspenderse mediante la adición de ácido fórmico al 0,1% y analizarse mediante CL-EM/EM.

2D-NanoCL-EM/EM: Todas las mezclas de péptidos etiquetados pueden separarse por medio de 2D-nanoCL en línea y analizarse mediante espectrometría de masas en tándem de electropulverización. Los experimentos pueden llevarse a cabo en un sistema 2D NanoCL Ultra de Eksigent conectado a un espectrómetro de masas LTQ Orbitrap Velos equipado con una fuente de iones de nanoelectropulverización (Thermo Electron, Bremen, Alemania).

Las mezclas de péptidos pueden inyectarse en una columna SCX de 5 cm (columna de polisulfoetil-aspartamida de 300 µm de DI, 5 µm, de PolyLC, Columbia, MD) con un flujo de 4 µl/min y eluirse en 10 segmentos de elución de intercambio iónico en una columna trampa C18 (2,5 cm, 100 µm de DI, 5 µm, 300 Å ProteoPep II de New Objective, Woburn, MA) y se lavó durante 5 minutos con H₂O/FA al 0,1%. La separación puede llevarse a cabo además a 300 nl/min usando un gradiente del 2-45% de B (H₂O/FA al 0,1% (disolvente A) y ACN/FA al 0,1% (disolvente B)) durante 120 minutos en una columna de sílice fundida de 15 cm (75 µm de DI, 5 µm, 300 Å ProteoPep II de New Objective, Woburn, MA).

Pueden adquirirse espectros de EM de exploración completa (m/z 300-2000) en el espectrómetro Orbitrap con una resolución de 30.000. Los iones más intensos (hasta 10) pueden aislarse secuencialmente para fragmentación usando la disociación en trampa C de alta energía (HCD) y excluir dinámicamente durante 30 segundos. Puede realizarse HCD con un ancho de aislamiento de 1,2 Da. Los iones de fragmentos resultantes pueden explorarse en el espectrómetro Orbitrap con una resolución de 7500. El espectrómetro LTQ Orbitrap Velos puede controlarse con Xcalibur 2.1 con Foundation 1.0.1.

Identificación y cuantificación de péptidos/proteínas: pueden identificarse péptidos y proteínas mediante la búsqueda automatizada de bases de datos usando el software Proteome Discoverer (Thermo Electron) con el motor de búsqueda Mascot frente a la base de datos SwissProt. Los parámetros de búsqueda pueden incluir 10 ppm para tolerancia a MS, 0,02 Da para tolerancia a EM2 y digestión completa con tripsina que permite hasta 2 escisiones perdidas. La carbamidometilación (C) puede establecerse tales como modificación fija. La oxidación (M), TMT6 y la desamidación (NQ) pueden establecerse como modificaciones dinámicas. Las identificaciones de péptidos y proteínas pueden filtrarse con el umbral significativo de Mascot ($p < 0,05$). A los filtros se les puede permitir un nivel de confianza de 99% de identificación de proteína (1% de FDA).

El software Proteome Discoverer puede aplicar factores de corrección en los iones indicadores y puede rechazar todos los valores de cuantificación si no están presentes todos los canales de cuantificación. La cuantificación relativa de proteína puede lograrse mediante la normalización a la intensidad media.

Con respecto al segundo tipo de datos, en algunas realizaciones a modo de ejemplo, la obtención de perfiles bioenergéticos de modelos normales y de cáncer pueden emplear el analizador Seahorse™ XF24 para permitir la comprensión de componentes de glicólisis y fosforilación oxidativa.

Específicamente, pueden sembrarse células en placas de cultivo de Seahorse a densidades óptimas. Estas células pueden sembrarse en placas en 100 μ l de medio o tratamiento y dejarse en un incubador a 37°C con el 5% de CO₂. Dos horas después, cuando las células se adhieren a la placa de 24 pocillos, pueden añadirse 150 μ l adicionales de o bien medio o bien disolución de tratamiento y las placas pueden dejarse en el incubador de cultivo durante la noche. Este procedimiento de siembra de dos etapas permite una distribución uniforme de las células en la placa de cultivo. Pueden hidratarse cartuchos de Seahorse que contienen el sensor de oxígeno y pH durante la noche en el fluido de calibración en un incubador sin CO₂ a 37°C. Tres fármacos mitocondriales se cargan normalmente en tres puertos en el cartucho. La oligomicina, un inhibidor del complejo III, FCCP, un desacoplador y rotenona, un inhibidor del complejo I pueden cargarse en los puertos A, B y C, respectivamente, del cartucho. Todos los fármacos de reserva pueden prepararse a una concentración de 10x en un medio DMEM no tamponado. Los cartuchos pueden incubarse en primer lugar con los compuestos mitocondriales en un incubador sin CO₂ durante aproximadamente 15 minutos antes del ensayo. Las placas de cultivo de Seahorse pueden lavarse en medios no tamponados basados en DMEM que contienen glucosa a una concentración que se encuentra en los medios de crecimiento normales. Las células pueden estratificarse con 630 μ l de los medios no tamponados y pueden equilibrarse en un incubador sin CO₂ antes de colocarse en el instrumento Seahorse con un cartucho precalibrado. El instrumento puede ejecutarse durante tres o cuatro bucles con un ciclo de mezcla, espera y medición para obtener una línea base, antes de que se inicie la inyección de fármacos a través del puerto. Puede haber dos bucles antes de que se introduzca el siguiente fármaco.

Los electrodos pueden registrar la TCO (tasa de consumo de oxígeno) y TAEC (tasa de acidificación extracelular) en una cámara de 7 μ l y pueden crearse con el cartucho empujando contra la placa de cultivo de Seahorse.

C. Integración de datos y generación de modelo *in silico*

Una vez que se han obtenido grupos de datos relevantes, la integración de los grupos de datos y la generación de modelos estadísticos implementados por ordenador pueden realizarse usando un sistema o plataforma de informática basada en IA (por ejemplo, la plataforma REFS™). Por ejemplo, un sistema basado en IA a modo de ejemplo puede producir redes basadas en simulación de asociaciones proteicas como controladores clave de puntos finales metabólicos (TAEC/TCO). Véase la figura 15. Algunos detalles de antecedentes relacionados con el sistema REFS™ pueden hallarse en Xing *et al.*, "Causal Modeling Using Network Ensemble Simulations of Genetic and Gene Expression Data Predicts Genes Involved in Rheumatoid Arthritis", PLoS Computational Biology, vol. 7, número 3, 1-19 (marzo de 2011) (e100105) y la patente estadounidense 7.512.497 concedida a Perival. En esencia, tal como se describió anteriormente, el sistema REFS™ es un sistema basado en IA que emplea algoritmos matemáticos para establecer relaciones causales entre las variables de entrada (por ejemplo, niveles de expresión de proteína, niveles de expresión de ARNm y los datos funcionales correspondientes, tales como valores de TCO/TAEC medidos en placas de cultivo de Seahorse). Este proceso se basa sólo en los datos de entrada, sin tener en cuenta el conocimiento previo existente sobre cualquier relación biológica potencial, establecida o verificada.

En particular, una ventaja significativa de la plataforma de la invención es que el sistema basado en IA se basa en los grupos de datos obtenidos del modelo celular, sin recurrir a ni tener en cuenta ningún conocimiento existente en la técnica referente al proceso biológico. Además, preferiblemente, no hay puntos de datos cortados estadística o artificialmente y, en cambio, todos los datos obtenidos se introducen en el sistema de IA para determinar las asociaciones proteicas. Por consiguiente, los modelos estadísticos resultantes generados a partir de la plataforma no tienen sesgo, puesto que no tienen en cuenta ninguna relación biológica conocida.

Específicamente, los datos de la proteómica y TAEC/TCO pueden introducirse en el sistema de información basado en IA, que construye modelos estadísticos basándose en asociaciones de datos, tal como se describió anteriormente. Las redes de asociaciones proteicas basadas en simulación se derivan entonces para cada enfermedad frente al escenario normal, incluyendo los tratamientos y estados que usan los siguientes métodos.

A continuación se muestra una descripción detallada de un proceso a modo de ejemplo para construir las redes generadas (por ejemplo, optimizadas o evolucionadas) con respecto a la figura 16. Tal como se describió anteriormente, los datos de la proteómica y los datos celulares funcionales se introducen en el sistema basado en IA (etapa 210). Los datos de entrada, que pueden ser datos sin procesar o datos mínimamente procesados, se someten a procesamiento previo, lo que puede incluir la normalización (por ejemplo, usando una función cuantil o patrones internos) (etapa 212). El procesamiento previo también puede incluir la imputación de valores de datos faltantes (por ejemplo, usando el algoritmo de los k vecinos más próximos (K-NN)) (etapa 212).

Los datos sometidos a procesamiento previo se usan para construir una biblioteca de fragmentos de red (etapa 214). Los fragmentos de red definen relaciones cuantitativas y continuas entre todos los posibles grupos pequeños (por ejemplo, grupos de 2-3 miembros o grupos de 2-4 miembros) de variables medidas (datos de entrada). Las relaciones entre las variables en un fragmento pueden ser lineales, logísticas, multinomiales, homocigotas dominantes o recesivas, etc. A la relación en cada fragmento se le asigna una puntuación probabilística bayesiana que refleja la probabilidad de que la relación candidata reciba los datos de entrada, y también penaliza la relación por su complejidad matemática. Al puntuar todas las posibles relaciones por parejas y a tres bandas (y en algunas realizaciones también relaciones a cuatro bandas) inferidas de los datos de entrada, pueden identificarse los fragmentos más probables en la biblioteca (los fragmentos probables). También se calculan parámetros cuantitativos de la relación basándose en los datos de entrada y se almacenan para cada fragmento. Pueden usarse diversos tipos de modelos en la enumeración de fragmentos incluyendo, pero sin limitarse a, regresión lineal, regresión logística, modelos ANOVA (análisis de varianza), modelos ANCOVA (análisis de covarianza), modelos de regresión no lineal/polinómica e incluso regresión no paramétrica. Las suposiciones anteriores sobre los parámetros del modelo pueden suponer distribuciones de Gull o penalizaciones de criterio de información bayesiano (CIB) relacionadas con el número de parámetros usados en el modelo. En un proceso de inferencia de red, cada red en un conjunto de redes de ensayo iniciales se construye a partir de un subgrupo de fragmentos en la biblioteca de fragmentos. Cada red de ensayo inicial en el conjunto de redes de ensayo iniciales se construye con un subgrupo diferente de los fragmentos de la biblioteca de fragmentos (etapa 216).

Una visión general de las representaciones matemáticas subyacentes a las redes bayesianas y los fragmentos de red, que se basa en Xing *et al.*, "Causal Modeling Using Network Ensemble Simulations of Genetic and Gene Expression Data Predicts Genes Involved in Rheumatoid Arthritis", PLoS Computational Biology, vol. 7, número 3, 1-19 (marzo de 2011) (e100105), se presenta a continuación.

Un sistema multivariante con variables aleatorias $X = X_1, \dots, X_n$ puede caracterizarse mediante una función de distribución de probabilidad multivariante $P(X_1, \dots, X_n; \Theta)$, que incluye un gran número de parámetros Θ . La función de distribución de probabilidad multivariante puede factorizarse y representarse mediante un producto de distribuciones de probabilidad condicional locales:

$$P(X_1, \dots, X_n; \Theta) = \prod_{i=1}^n P_i(X_i | Y_{j_1}, \dots, Y_{j_{K_i}}; \Theta_i),$$

en el que cada variable X_i es independiente de sus variables no descendientes dadas sus variables parentales K_i , que son $Y_{j_1}, \dots, Y_{j_{K_i}}$. Después de la factorización, cada distribución de probabilidad local tiene sus propios parámetros Θ_i .

La función de distribución de probabilidad multivariante puede factorizarse de diferentes maneras con cada factorización particular y siendo los parámetros correspondientes un modelo probabilístico distinto. Cada (modelo de) factorización particular puede representarse mediante un grafo acrílico dirigido (GAD) que tiene un vértice para cada variable X_i y aristas dirigidas entre vértices que representan dependencias entre variables en las distribuciones

condicionales locales $P_i(X_i | Y_{j_1}, \dots, Y_{j_{K_i}})$. Los subgrafos de un GAD, cada uno con un vértice y aristas dirigidas asociadas son fragmentos de red.

Un modelo evoluciona o se optimiza determinando la factorización más probable y los parámetros más probables dados los datos de entrada. Esto puede describirse como "aprender una red bayesiana" o, dicho de otro modo, dado un grupo de entrenamiento de datos de entrada, hallar una red que corresponda de la mejor manera a los datos de entrada. Esto se logra mediante el uso de una función de puntuación que evalúa cada red con respecto a los datos de entrada.

Se usa un marco bayesiano para determinar la probabilidad de una factorización dados los datos de entrada. La ley de Bayes establece que la probabilidad *a posteriori*, $P(d|M)$, de un modelo M, dado que los datos D son proporcionales al producto de la probabilidad *a posteriori* de los datos dadas las suposiciones del modelo, $P(d|M)$, multiplicado por la probabilidad *a priori* del modelo, $P(M)$, suponiendo que la probabilidad de los datos, $P(d)$, es constante en todos los modelos. Esto se expresa en la siguiente ecuación:

$$P(M|D) = \frac{P(D|M) * P(M)}{P(D)}$$

5 La probabilidad *a posteriori* de que los datos adopten el modelo es la integral de la probabilidad de datos para la distribución previa de parámetros:

$$P(D|M) = \int P(D|M(\Theta))P(\Theta|M)d\Theta$$

10 Suponiendo que todos los modelos son igualmente probables (es decir, que $P(M)$ es una constante), la probabilidad *a posteriori* del modelo M dados los datos D puede factorizarse en el producto de integrales para los parámetros para cada fragmento de red local M_i tal como sigue:

$$P(M|D) = \prod_{i=1}^n \int P_i(X_i | Y_{j1}, \dots, Y_{jk_i}; \Theta_i)$$

15 Obsérvese que en la ecuación anterior, se ha omitido un término constante principal. En algunas realizaciones, un criterio de información bayesiano (CIB), que toma un logaritmo negativo de la probabilidad *a posteriori* del modelo $P(d|M)$ puede usarse para “puntuar” cada modelo de la siguiente manera:

$$S_{tot}(M) = -\log P(M|D) = \sum_{i=1}^n S(M_i),$$

20 donde la puntuación total S_{tot} para un modelo M es una suma de las puntuaciones locales S_i para cada fragmento de red local. El CIB proporciona además una expresión para determinar una puntuación de cada fragmento de red individual:

$$25 S(M_i) \approx S_{BIC}(M_i) = S_{MLE}(M_i) + \frac{\kappa(M_i)}{2} \log N$$

30 dónde $\kappa(M_i)$ es el número de parámetro de ajuste en el modelo M_i y N es el número de muestras (puntos de datos). $S_{MLE}(M_i)$ es el logaritmo negativo de la función de probabilidad para un fragmento de red, que puede calcularse a partir de las relaciones funcionales usadas para cada fragmento de red. Para una puntuación de CIB, cuanto menor sea la puntuación, más probable es que un modelo se ajuste a los datos de entrada.

35 El conjunto de redes de ensayo está optimizado globalmente, lo que puede describirse como optimización o evolución de las redes (etapa 218). Por ejemplo, las redes de ensayo pueden evolucionar y optimizarse según un algoritmo de muestreo de Monte Carlo de Metropolis. La hibridación simulada puede usarse para optimizar o evolucionar cada red de ensayo en el conjunto a través de transformaciones locales. En un ejemplo de procesos de hibridación simulados, cada red de ensayo se cambia al añadir un fragmento de red de la biblioteca, al deletar un fragmento de red de la red de ensayo, al sustituir un fragmento de red o al cambiar la topología de la red, y luego se calcula una nueva puntuación para la red. En términos generales, si la puntuación mejora, el cambio se mantiene y si la puntuación empeora, el cambio se rechaza. Un parámetro de “temperatura” permite algunos cambios locales que empeoran la puntuación que va a mantenerse, lo que ayuda al proceso de optimización para evitar algunos mínimos locales. El parámetro “temperatura” se reduce a lo largo del tiempo para permitir que el proceso de optimización/evolución converja.

45 La totalidad o parte del proceso de inferencia de red puede realizarse en paralelo para las diferentes redes de ensayo. Cada red puede optimizarse en paralelo en un procesador independiente y/o en un dispositivo de computación separado. En algunas realizaciones, el proceso de optimización puede llevarse a cabo en una superordenador que incorpore de cientos a miles de procesadores que funcionan en paralelo. La información puede compartirse entre los procesos de optimización realizados en procesadores paralelos.

50 El proceso de optimización puede incluir un filtro de red que elimina cualquier red del conjunto que no cumpla con un criterio de umbral para la puntuación general. La red eliminada puede reemplazarse por una nueva red inicial. Además, cualquier red que no sea “sin escala” puede eliminarse del conjunto. Después de que el conjunto de redes se ha optimizado o evolucionado, el resultado puede denominarse un conjunto de redes de modelo celular generadas, que pueden denominarse colectivamente red de consenso generada.

55 D. Simulación para extraer información de relación cuantitativa y para predicción

La simulación puede usarse para extraer información de parámetros cuantitativos con respecto a cada relación en

5 las redes de modelo celular generadas (etapa 220). Por ejemplo, la simulación para la extracción de información cuantitativa puede implicar perturbar (aumentar o disminuir) cada nodo en la red en 10 veces y calcular las distribuciones *a posteriori* para los otros nodos (por ejemplo, proteínas) en los modelos. Los puntos finales se comparan mediante la prueba de la t con la suposición de 100 muestras por grupo y el punto de corte de significación de 0,01. La estadística de la prueba de la t es la mediana de 100 pruebas de la t. Mediante el uso de esta técnica de simulación, se genera un AUC (área bajo la curva) que representa la intensidad de predicción y el cambio en veces que representa la magnitud *in silico* de un nodo que conduce a un punto final para cada relación en el conjunto de redes.

10 Puede emplearse un módulo de cuantificación de relaciones de un sistema informático local para dirigir el sistema basado en IA para realizar las perturbaciones y extraer la información de AUC y la información de veces. La información cuantitativa extraída puede incluir cambio en veces y AUC para cada arista que conecta una nota principal a un nodo secundario. En algunas realizaciones, puede usarse un programa R personalizado para extraer la información cuantitativa.

15 En algunas realizaciones, el conjunto de redes de modelo celular generadas puede usarse a través de la simulación para predecir respuestas a cambios en las condiciones, que luego pueden verificarse a través de experimentos basados en células o basados en animales, de laboratorio de investigación.

20 La salida del sistema basado en IA pueden ser parámetros de relación cuantitativos y/u otras predicciones de simulación (222).

E. Generación de redes diferenciales (Delta)

25 Puede usarse un módulo de creación de red diferencial para generar redes diferenciales (delta) entre redes de modelo celular generadas y redes de modelo celular de comparación generadas. Tal como se describió anteriormente, en algunas realizaciones, la red diferencial compara todos los parámetros cuantitativos de las relaciones en las redes del modelo celular generado y la red del modelo celular comparativo generada. Los parámetros cuantitativos para cada relación en la red diferencial se basan en la comparación. En algunas realizaciones, puede realizarse un diferencial entre diversas redes diferenciales, que puede denominarse una red delta-delta. A continuación se describe un ejemplo de una red delta-delta con respecto a la figura 26 en la sección de ejemplos. El módulo de creación de red diferencial puede ser un programa o *script* escrito en PERL.

F. Visualización de redes

35 Los valores de relación para el conjunto de redes y para las redes diferenciales pueden visualizarse usando un programa de visualización de red (por ejemplo, la plataforma de código abierto Cytoscape para análisis y visualización de redes complejas del consorcio Cytoscape). En las representaciones visuales de las redes, el grosor de cada arista (por ejemplo, cada línea que conecta las proteínas) representa la intensidad del cambio en veces. Las aristas también son direccionales, lo que indica causalidad, y cada arista tiene un nivel de confianza de predicción asociado.

G. Sistema informático a modo de ejemplo

45 La figura 17 representa esquemáticamente un sistema/entorno informático a modo de ejemplo que puede emplearse en algunas realizaciones para comunicarse con el sistema de informática basado en IA, para generar redes diferenciales, para visualizar redes, para guardar y almacenar datos, y/o para interactuar con un usuario. Tal como se explicó anteriormente, los cálculos para un sistema de informática basado en IA pueden realizarse en un superordenador independiente con cientos o miles de procesadores paralelos que interactúan, directa o indirectamente, con el sistema informático a modo de ejemplo. El entorno incluye un dispositivo 100 de computación con dispositivos periféricos asociados. El dispositivo 100 de computación es programable para implementar el código 150 ejecutable para realizar diversos métodos, o partes de métodos, que se enseñan en el presente documento. El dispositivo 100 de computación incluye un dispositivo 116 de almacenamiento, tal como un disco duro, un CD-ROM u otro medio legible por ordenador no transitorio. El dispositivo 116 de almacenamiento puede almacenar un sistema 118 operativo y otro software relacionado. El dispositivo 100 de computación puede incluir además la memoria 106. La memoria 106 puede comprender una memoria de sistema informático o memoria de acceso aleatorio, tal como DRAM, SRAM, EDO RAM, etc. La memoria 106 puede comprender también otros tipos de memoria, o combinaciones de los mismos. El dispositivo 100 de computación puede almacenar, en el dispositivo 116 de almacenamiento y/o la memoria 106, instrucciones para implementar y procesar cada parte del código 150 ejecutable.

60 El código 150 ejecutable puede incluir código para comunicarse con el sistema 190 de informática basado en IA, para generar redes diferenciales (por ejemplo, un módulo de creación de red diferencial), para extraer información de relación cuantitativa del sistema de informática basado en IA (por ejemplo, un módulo de cuantificación de relación) y para visualizar redes (por ejemplo, Cytoscape).

65

En algunas realizaciones, el dispositivo 100 de computación puede comunicarse directa o indirectamente con el sistema 190 de informática basado en IA (por ejemplo, un sistema para ejecutar REFS). Por ejemplo, el dispositivo 100 de computación puede comunicarse con el sistema 190 de informática basado en IA transfiriendo archivos de datos (por ejemplo, tramas de datos) al sistema 190 informático basado en IA a través de una red. Además, el dispositivo 100 de computación puede tener un código 150 ejecutable que proporciona una interfaz e instrucciones para el sistema 190 de informática basado en IA.

En algunas realizaciones, el dispositivo 100 de computación puede comunicarse directa o indirectamente con uno o más sistemas 180 experimentales que proporcionan datos para el grupo de datos de entrada. Los sistemas 180 experimentales para generar datos pueden incluir sistemas para proteómica basados en espectrometría de masas, expresión génica de microalineamientos, expresión génica de qPCR, metabólica datos en espectrometría de masas y lipídica datos en espectrometría de masas, microalineamientos de SNP, un panel de ensayos funcionales y otras plataformas y tecnologías de biología *in vitro*.

El dispositivo 100 de computación también incluye el procesador 102, y puede incluir uno o más procesadores adicionales 102', para ejecutar software almacenado en la memoria 106 y otros programas para controlar hardware del sistema, dispositivos periféricos y/o hardware periférico. El procesador 102 y el/los procesador(es) 102' pueden ser un procesador de un único núcleo o un procesador de múltiples núcleos (104 y 104'). La virtualización puede emplearse en el dispositivo 100 de computación de modo que la infraestructura y los recursos en el dispositivo de computación puedan compartirse dinámicamente. También pueden usarse procesadores virtualizados con el código 150 ejecutable y otro software en el dispositivo 116 de almacenamiento. Puede proporcionarse una máquina 114 virtual para manejar un proceso que se ejecuta en múltiples procesadores, de modo que el proceso parece estar usando sólo un recurso de computación en lugar de múltiples. También pueden usarse diversas máquinas virtuales con un procesador.

Un usuario puede interactuar con el dispositivo 100 de computación a través de un dispositivo 122 de presentación visual, tal como un monitor de ordenador, que puede mostrar una interfaz 124 de usuario o cualquier otra interfaz. La interfaz 124 de usuario del dispositivo 122 de presentación visual puede usarse para presentar visualmente datos sin procesar, representaciones visuales de redes, etc. El dispositivo 122 de presentación visual también puede presentar visualmente otros aspectos o elementos de realizaciones a modo de ejemplo (por ejemplo, un icono para el dispositivo 116 de almacenamiento). El dispositivo 100 de computación puede incluir otros dispositivos de I/O (E/S) tales como un teclado o una interfaz 108 táctil multipunto (por ejemplo, una pantalla táctil) y un dispositivo 110 señalador, (por ejemplo, un ratón, una bola de seguimiento y/o un panel táctil) para recibir la entrada de un usuario. El teclado 108 y el dispositivo 110 señalador pueden estar conectados al dispositivo 122 de presentación visual y/o al dispositivo 100 de computación a través de una conexión por cable y/o inalámbrica.

El dispositivo 100 de computación puede incluir una interfaz 112 de red para interactuar con un dispositivo 126 de red a través de una red de área local (LAN), una red de área amplia (WAN) o Internet a través de una variedad de conexiones incluyendo, pero sin limitarse a, líneas telefónicas convencionales, enlaces LAN o WAN (por ejemplo, 802.11, T1, T3, 56kb, X.25), conexiones de banda ancha (por ejemplo, ISDN, Frame Relay, ATM), conexiones inalámbricas, red de área del controlador (CAN), o alguna combinación de cualquiera o todos los anteriores. La interfaz 112 de red puede comprender un adaptador de red incorporado, tarjeta de interfaz de red, tarjeta de red PCMCIA, adaptador de red de bus de tarjeta, adaptador de red inalámbrico, adaptador de red USB, módem o cualquier otro dispositivo adecuado para permitir que el dispositivo 100 de computación se interconecte con cualquier tipo de red capaz de comunicarse y realizar las operaciones descritas en el presente documento.

Además, el dispositivo 100 de computación puede ser cualquier sistema informático, tal como una estación de trabajo, ordenador de escritorio, servidor, ordenador portátil, ordenador de mano u otra forma de dispositivo de computación o de telecomunicaciones que sea capaz de comunicarse y que tenga suficiente potencia de procesador y capacidad de memoria para realizar las operaciones descritas en el presente documento.

El dispositivo 100 de computación puede ejecutar cualquier sistema 118 operativo, como cualquiera de las versiones de los sistemas operativos MICROSOFT WINDOWS, las diferentes versiones de los sistemas operativos Unix y Linux, cualquier versión de MACOS para ordenadores Macintosh, cualquier sistema operativo incorporado, cualquier sistema operativo en tiempo real, cualquier sistema operativo de código abierto, cualquier sistema operativo patentado, cualquier sistema operativo para dispositivos de computación móviles o cualquier otro sistema operativo capaz de ejecutarse en el dispositivo de computación y realizar las operaciones descritas en el presente documento. El sistema operativo puede estar ejecutándose en modo nativo o en modo emulado.

IV. Modelos para un sistema biológico y usos de los mismos

A. Establecer un modelo para un sistema biológico

Prácticamente todos los sistemas o procesos biológicos implican interacciones complicadas entre diferentes tipos de células y/o sistemas de órganos. La perturbación de funciones críticas en un tipo de célula u órgano puede conducir a efectos secundarios en otros tipos de células y órganos que interactúan, y tales cambios posteriores pueden a

su vez responder a los cambios iniciales y provocar complicaciones adicionales. Por tanto, es beneficioso diseccionar un sistema o proceso biológico dado en sus componentes, tales como interacción entre pares de tipos de células u órganos, y sondear sistemáticamente las interacciones entre estos componentes para obtener una visión más completa y global del sistema o proceso biológico.

Por consiguiente, la presente invención proporciona modelos celulares para sistemas biológicos. Con este propósito, los solicitantes han construido modelos celulares para diversos sistemas biológicos a modo de ejemplo que se han empleado en la tecnología de plataforma de descubrimiento objeto. Los solicitantes han realizado experimentos con los modelos celulares usando la tecnología de plataforma de descubrimiento objeto para generar redes de relaciones causales de consenso, incluyendo las relaciones causales únicas en el sistema biológico, y así identificar “moduladores” o “controladores” moleculares críticos importantes para los sistemas o procesos biológicos particulares.

Una ventaja significativa de la tecnología de plataforma y sus componentes, por ejemplo, los modelos celulares personalizados y los grupos de datos obtenidos de los modelos celulares, es que una red de relaciones causales de consenso inicial de “primera generación” generada para un sistema o proceso biológico puede evolucionar de manera continua o expandirse a lo largo del tiempo, por ejemplo, mediante la introducción de líneas/tipos de células adicionales y/o condiciones adicionales. Pueden recopilarse datos adicionales del modelo celular evolucionado, es decir, datos de la(s) parte(s) recién añadida(s) del modelo celular. Los nuevos datos recopilados de un modelo celular expandido o evolucionado, es decir, de la(s) parte(s) recién añadida(s) del modelo celular, pueden introducirse luego en los grupos de datos usados previamente para generar la red de relaciones causales de consenso de “primera generación” con el fin de generar una red de relaciones causales de consenso de “segunda generación” más robusta. Pueden identificarse nuevas relaciones causales únicas del sistema biológico a partir de la red de relaciones causales de consenso de “segunda generación”. De esta manera, la evolución del modelo celular proporciona una evolución de las redes de relaciones causales de consenso, proporcionando así conocimientos nuevos y/o más fiables sobre los moduladores del sistema biológico. De esta manera, tanto los modelos celulares, los grupos de datos de los modelos celulares como las redes de relaciones causales generadas a partir de los modelos celulares usando los métodos de la tecnología de plataforma pueden evolucionar constantemente y sumarse al conocimiento previo obtenido de la tecnología de plataforma.

Por consiguiente, la invención proporciona redes de relaciones causales de consenso generadas a partir de los modelos celulares empleados en la tecnología de plataforma. Estas redes de relaciones causales de consenso pueden ser redes de relaciones causales de consenso de primera generación, o pueden ser redes de relaciones causales de consenso de generación múltiple, por ejemplo, redes de relaciones causales de consenso de 2^a, 3^a, 4^a, 5^a, 6^a, 7^a, 8^a, 9^a, 10^a, 11^a, 12^a, 13^a, 14^a, 15^a, 16^a, 17^a, 18^a, 19^a, 20^a generación o mayor. Además, la invención proporciona redes de relaciones causales de consenso simuladas generadas a partir de los modelos celulares empleados en la tecnología de plataforma. Estas redes de relaciones causales de consenso simuladas pueden ser redes de relaciones causales de consenso simuladas de primera generación, o pueden ser redes de relaciones causales de consenso simuladas de generación múltiple, por ejemplo, redes de relaciones causales de consenso simuladas de 2^a, 3^a, 4^a, 5^a, 6^a, 7^a, 8^a, 9^a, 10^a, 11^a, 12^a, 13^a, 14^a, 15^a, 16^a, 17^a, 18^a, 19^a, 20^a generación o mayor. La invención proporciona además redes delta y redes delta-delta generadas a partir de cualquiera de las redes de relaciones causales de consenso de la invención.

Un modelo celular personalizado para un sistema o proceso biológico comprende una o más células asociadas con el sistema biológico. El modelo para un sistema/proceso biológico puede establecerse para simular un entorno de sistema biológico, por ejemplo, el entorno de una célula cancerosa *in vivo*, creando condiciones (por ejemplo, condiciones de cultivo celular) que imitan un aspecto característico del sistema o proceso biológico.

Pueden incluirse múltiples células del mismo o diferente origen, en oposición a un único tipo de célula, en el modelo celular. En una realización, se incluyen al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50 o más líneas celulares o tipos de células diferentes en el modelo celular. En una realización, las células son todas del mismo tipo, por ejemplo, todas las células de cáncer de mama o células vegetales, pero son diferentes líneas celulares establecidas, por ejemplo, diferentes líneas celulares establecidas de células de cáncer de mama o células vegetales. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 líneas celulares o tipos de células diferentes.

Los ejemplos de tipos de células que pueden incluirse en los modelos celulares de la invención incluyen, sin limitación, células humanas, células animales, células de mamíferos, células vegetales, levaduras, bacterias u hongos. En una realización, las células del modelo celular pueden incluir células enfermas, tales como células cancerosas o células infectadas por bacterias o virus. En una realización, las células del modelo celular pueden incluir células asociadas a enfermedad, tales como células implicadas en diabetes, obesidad o estado de enfermedad cardiovascular, por ejemplo, células de músculo liso aórtico o hepatocitos. El experto reconocerá aquellas células que están implicadas o asociadas con un estado/proceso biológico particular, por ejemplo, estado/proceso de enfermedad, y cualquiera de estas células puede incluirse en un modelo celular de la invención.

Los modelos celulares de la invención pueden incluir una o más “células de control”. En una realización, una célula de control puede ser una célula no tratada o no perturbada. En otra realización, una “célula de control” puede ser una célula normal, por ejemplo, no enferma. En una realización, se incluyen al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 o más células de control diferentes en el modelo celular. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 líneas celulares de control o tipos de células de control diferentes. En una realización, las células de control son todas del mismo tipo pero son diferentes líneas celulares establecidas de ese tipo de células. En una realización, como control, una o más líneas celulares normales, por ejemplo, no enfermas, se cultivan en condiciones similares, y/o se exponen a la misma perturbación que las células primarias del modelo celular para identificar proteínas o rutas únicas para el estado o proceso biológico.

Un modelo celular personalizado de la invención también puede comprender condiciones que imitan un aspecto característico del estado o proceso biológico. Por ejemplo, pueden seleccionarse condiciones de cultivo celular que se aproximen estrechamente a las condiciones de una célula cancerosa en un entorno tumoral *in vivo*, o de una célula de músculo liso aórtico de un paciente que padece una enfermedad cardiovascular. En algunos casos, las condiciones son condiciones de estrés. Pueden emplearse diversas condiciones/ factores estresantes en los modelos celulares de la invención. En una realización, estas condiciones/ factores estresantes pueden constituir la “perturbación”, por ejemplo, estímulo externo, para los sistemas celulares. Una condición de estrés a modo de ejemplo es la hipoxia, una condición que se encuentra normalmente, por ejemplo, dentro de tumores sólidos. Puede inducirse hipoxia usando métodos reconocidos en la técnica. Por ejemplo, puede inducirse hipoxia colocando sistemas celulares en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que puede inundarse con una mezcla de gases industriales que contiene el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Pueden medirse los efectos después de un periodo predeterminado, por ejemplo, a las 24 horas después del tratamiento de hipoxia, con y sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM). Asimismo, el tratamiento con ácido láctico imita un entorno celular en el que la actividad de glicólisis es alta. El estrés inducido por ácido láctico puede investigarse a una concentración de ácido láctico final de aproximadamente 12,5 mM en un tiempo predeterminado, por ejemplo, a las 24 horas, con o sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM). La hiperglucemia es una condición que se encuentra tanto en la diabetes como en el cáncer. Una condición hiperglucémica típica que puede usarse para tratar las células en cuestión incluye glucosa de grado de cultivo al 10% añadida a medios adecuados para elevar la concentración final de glucosa en los medios hasta aproximadamente 22 mM. La hiperlipidemia es una condición que se encuentra, por ejemplo, en la obesidad y la enfermedad cardiovascular. Las condiciones hiperlipidémicas pueden proporcionarse cultivando células en medios que contienen palmitato de sodio 0,15 mM. La hiperinsulinemia es una condición que se encuentra, por ejemplo, en la diabetes. Las condiciones hiperinsulinémicas pueden inducirse mediante el cultivo de las células en medios que contienen insulina 1000 nM.

Las condiciones individuales pueden investigarse por separado en los modelos de células personalizadas de la invención, y/o pueden combinarse entre sí. En una realización, una combinación de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más condiciones que reflejan o simulan diferentes aspectos característicos del sistema biológico se investigan en el modelo celular personalizado. En una realización, condiciones individuales y, además, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50 o más de las condiciones que reflejan o simulan diferentes aspectos característicos del sistema biológico se investigan en el modelo celular personalizado. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 ó 10 y 50 condiciones diferentes.

Una vez que se construye el modelo celular personalizado, pueden aplicarse una o más “perturbaciones” al sistema, tales como variación genética de un paciente a otro, o con/sin tratamiento con determinados fármacos o profármacos. Véase la figura 15D. Los efectos de tales perturbaciones en el sistema del modelo celular pueden medirse usando diversos medios reconocidos en la técnica o patentados, tal como se describe en la sección III.B a continuación.

El modelo celular personalizado puede exponerse a una perturbación, por ejemplo, una “perturbación ambiental” o “componente de estímulo externo”. La “perturbación ambiental” o el “componente de estímulo externo” pueden ser endógenos con respecto al entorno celular (por ejemplo, el entorno celular contiene ciertos niveles del estimulante, y se añade más cantidad del mismo para aumentar su nivel), o puede ser exógeno con respecto al entorno celular (por ejemplo, el estimulante/perturbación está ausente en gran medida del entorno celular antes de la alteración). El entorno celular puede verse alterado adicionalmente mediante cambios secundarios que resultan de añadir la perturbación ambiental o el componente de estímulo externo, puesto que el componente de estímulo externo puede cambiar la salida celular del sistema celular, incluyendo moléculas secretadas en el entorno celular por el sistema celular. La perturbación ambiental o el componente de estímulo externo puede incluir cualquier estímulo físico y/o químico externo que pueda afectar a la función celular. Esto puede incluir cualquier molécula orgánica o inorgánica grande o pequeña, productos químicos naturales o sintéticos, cambio de temperatura, cambio de pH, radiación, luz (UVA, UVB, etc.), microondas, ondas sónicas, corriente eléctrica, campos magnéticos modulados o no modulados, etc. La perturbación ambiental o el componente de estímulo externo también puede incluir una

modificación o mutación genética introducida o un vehículo (por ejemplo, un vector) que provoca una modificación/mutación genética.

(i) Sistemas celulares de interferencia

5 En determinadas situaciones, cuando se desea investigar la interacción entre dos o más sistemas celulares, puede formarse un “sistema celular de interferencia”, por ejemplo, poniendo en contacto el entorno celular modificado de un primer sistema celular con un segundo sistema celular para afectar a la salida celular del segundo sistema celular.

10 Tal como se usa en el presente documento, “sistema celular de interferencia” comprende dos o más sistemas celulares, en los que el entorno celular de al menos un sistema celular entra en contacto con un segundo sistema celular, de modo que al menos una salida celular en el segundo sistema celular se cambia o ve afectada. En determinadas realizaciones, los sistemas celulares dentro del sistema celular de interferencia pueden estar en contacto directo entre sí. En otras realizaciones, ninguno de los sistemas celulares está en contacto directo entre sí.

15 Por ejemplo, en determinadas realizaciones, el sistema celular de interferencia puede tener la forma de una placa Transwell, en la que un primer sistema celular está creciendo en un inserto y un segundo sistema celular está creciendo en un compartimento de pocillo correspondiente. Los dos sistemas celulares pueden estar en contacto con medios iguales o diferentes, y pueden intercambiar parte o la totalidad de los componentes de los medios. El componente de estímulo externo añadido a un sistema celular puede absorberse sustancialmente por un sistema celular y/o degradarse antes de que tenga la posibilidad de difundir al otro sistema celular. Alternativamente, el componente de estímulo externo puede aproximarse o alcanzar eventualmente un equilibrio dentro de los dos sistemas celulares.

20 En determinadas realizaciones, el sistema celular de interferencia puede adoptar la forma de sistemas celulares cultivados por separado, en el que cada sistema celular puede tener su propio medio y/o condiciones de cultivo (temperatura, contenido de CO₂, pH, etc.), o condiciones de cultivo similares o idénticas. Los dos sistemas celulares pueden entrar en contacto, por ejemplo, tomando el medio condicionado de un sistema celular y poniéndolo en contacto con otro sistema celular. También pueden efectuarse contactos intercelulares directos entre los dos sistemas celulares, si se desea. Por ejemplo, las células de los dos sistemas celulares pueden cocultivarse en cualquier punto, si se desea, y los sistemas celulares cocultivados pueden separarse más tarde, por ejemplo, mediante clasificación FACS cuando las células en al menos un sistema celular tienen un marcador o una etiqueta clasificable (tal como una proteína marcadora fluorescente GFP expresada de manera estable).

25 De manera similar, en determinadas realizaciones, el sistema celular de interferencia puede ser simplemente un cocultivo. El tratamiento selectivo de las células en un sistema celular puede efectuarse tratando en primer lugar las células en ese sistema celular, antes de cultivar las células tratadas en cocultivo con células en otro sistema celular. La configuración del sistema celular de interferencia de cocultivo puede ser útil cuando se desea estudiar, por ejemplo, los efectos sobre un segundo sistema celular provocados por cambios en la superficie celular en un primer sistema celular, después de la estimulación del primer sistema celular mediante un componente de estímulo externo.

30 El sistema celular de interferencia de la invención es particularmente adecuado para explorar el efecto de determinado componente de estímulo externo predeterminado sobre la salida celular de uno o ambos sistemas celulares. El efecto primario de tal estímulo sobre el primer sistema celular (con el que el estímulo entra en contacto directamente) puede determinarse comparando las salidas celulares (por ejemplo, nivel de expresión de proteína) antes y después del primer contacto del sistema celular con el estímulo externo que, tal como se usa en el presente documento, pueden denominarse “diferenciales de salida celular (significativos)”. El efecto secundario de tal estímulo sobre el segundo sistema celular, que está mediado a través del entorno celular modificado del primer sistema celular (tal como su secretoma), también puede medirse de manera similar. En ese caso, puede realizarse una comparación, por ejemplo, en el proteoma del segundo sistema celular entre el proteoma del segundo sistema celular *con* el tratamiento con estímulo externo en el primer sistema celular y el proteoma del segundo sistema celular *sin* el tratamiento con estímulo externo en el primer sistema celular. Cualquier cambio significativo observado (en el proteoma o en cualquier otra salida celular de interés) puede denominarse un “diferencial significativo de interferencia celular”.

35 Al realizar mediciones de salida celular (tales como expresión de proteína), puede usarse o bien la cantidad de expresión absoluta o bien el nivel de expresión relativo. Por ejemplo, para determinar el nivel de expresión relativo de proteína de un segundo sistema celular, la cantidad de cualquier proteína dada en el segundo sistema celular, con o sin el estímulo externo al primer sistema celular, puede compararse con una línea celular de control adecuada y mezcla de líneas celulares y dado un valor de aumento en veces o disminución en veces. Un nivel umbral predeterminado para tal aumento en veces (por ejemplo, un aumento de al menos 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2, 2,5, 3, 3,5, 4, 4,5, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 ó 100 veces o más) o disminución en veces (por ejemplo, al menos una disminución de hasta 0,95, 0,9, 0,8, 0,75, 0,7, 0,6, 0,5, 0,45, 0,4, 0,35, 0,3, 0,25, 0,2, 0,15, 0,1 o 0,05 veces, o el 90%, el 80%, el 75%, el 70%, el 65%, el 60%, el 55%, el 50%, el 45%, el 40%, el 35%, el 30%, el 25%, el 20%, el 15%, el 10% o el 5% o menos) pueden usarse para seleccionar diferenciales de interferencia celular. Todos los valores presentados en la lista anterior también pueden ser el límite superior o

inferior de intervalos, por ejemplo, entre 1,5 y 5 veces, entre 2 y 10 veces, entre 1 y 2 veces, o entre 0,9 y 0,7 veces, que pretenden formar parte de esta invención.

5 En la totalidad de la presente solicitud, todos los valores presentados en una lista, por ejemplo, tales como los anteriores, también pueden ser el límite superior o inferior de intervalos que pretenden formar parte de esta invención.

10 Para ilustrar, en un sistema de dos células a modo de ejemplo establecido para imitar aspectos de un modelo de enfermedad cardiovascular, una línea celular de músculo liso cardiaco (primer sistema celular) puede tratarse con una condición de hipoxia (un componente de estímulo externo) y cambios de proteoma en una línea celular de riñón (segundo sistema celular) resultante del contacto de las células renales con el medio condicionado del músculo liso cardiaco puede medirse usando espectrometría de masas cuantitativa convencional. Pueden determinarse 15 diferenciales significativos de interferencia celular en estas células renales, basándose en la comparación con un control apropiado (por ejemplo, células de riñón cultivadas de manera similar puestas en contacto con medio condicionado de células de músculo liso cardiaco cultivadas de manera similar no tratadas con condiciones de hipoxia).

20 No todos los diferenciales significativos de interferencia celular observados pueden tener significación biológica. Con respecto a cualquier sistema biológico dado para el que se aplica la evaluación biológica por interrogación objeto, algunos (o tal vez todos) de los diferenciales significativos de interferencia celular pueden ser “determinantes” con respecto al problema biológico específico en cuestión, por ejemplo, o bien responsable de provocar una enfermedad (una diana potencial para intervención terapéutica) o bien es un biomarcador para el estado de enfermedad (un posible factor de diagnóstico o pronóstico).

25 Dichos diferenciales de interferencia determinantes pueden seleccionarse por un usuario final del método objeto, o pueden seleccionarse por un programa de software de bioinformática, tal como programa de análisis de rutas comparativo habilitado por DAVID o el programa de análisis de rutas KEGG. En determinadas realizaciones, se usa más de un programa de software de bioinformática, y se prefieren los resultados de consenso de dos o más programas de software de bioinformática.

30 Tal como se usa en el presente documento, los “diferenciales” de salidas celulares incluyen diferencias (por ejemplo, niveles aumentados o disminuidos) en uno cualquiera o más parámetros de las salidas celulares. Por ejemplo, en cuanto al nivel de expresión de proteína, los diferenciales entre dos salidas celulares, tales como las salidas asociadas con un sistema celular antes y después del tratamiento con un componente de estímulo externo, pueden medirse y cuantificarse mediante el uso de tecnologías reconocidas en la técnica, tales como ensayos basados en 35 espectrometría de masas (por ejemplo, iTRAQ, 2D-CL-EM/EM, etc.).

(ii) Modelos específicos de cáncer

40 Un ejemplo de un sistema o proceso biológico es el cáncer. Como cualquier otro proceso o sistema biológico complicado, el cáncer es un estado patológico complicado caracterizado por múltiples aspectos únicos. Por ejemplo, debido a su alta velocidad de crecimiento, muchas células cancerosas están adaptadas para crecer en condiciones de hipoxia, tienen glicólisis regulada por incremento y rutas metabólicas de fosforilación oxidativa reducidas. Como resultado, las células cancerosas pueden reaccionar de manera diferente a una perturbación ambiental, tal como 45 tratamiento con un fármaco potencial, en comparación con la reacción de una célula normal en respuesta al mismo tratamiento. Por tanto, sería interesante descifrar las respuestas únicas del cáncer al tratamiento farmacológico en comparación con las respuestas de las células normales. Con este propósito, puede establecerse un modelo de cáncer personalizado para simular el entorno de una célula cancerosa, por ejemplo, dentro de un tumor *in vivo*, eligiendo líneas celulares de cáncer apropiadas y creando condiciones de cultivo celular que imiten un aspecto 50 característico del estado o proceso de enfermedad. Por ejemplo, pueden seleccionarse condiciones de cultivo celular que se aproximen estrechamente a las condiciones de una célula cancerosa en un tumor *in vivo*, o para imitar diversos aspectos del crecimiento del cáncer, aislando diferentes condiciones de crecimiento de las células cancerosas.

55 Pueden incluirse en el modelo de cáncer múltiples células cancerosas del mismo o diferente origen (por ejemplo, líneas de cáncer PaCa2, HepG2, PC3 y MCF7), en oposición a un único tipo de célula cancerosa. En una realización, se incluyen al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 o más líneas celulares de cáncer o tipos de células cancerosas diferentes en el modelo de cáncer. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 líneas o tipos de células cancerosas diferentes.

60 En una realización, las células cancerosas son todas del mismo tipo, por ejemplo, todas las células de cáncer de mama, pero son diferentes líneas celulares establecidas, por ejemplo, diferentes líneas celulares establecidas de 65 cáncer de mama.

Los ejemplos de tipos de células cancerosas que pueden incluirse en el modelo de cáncer incluyen, sin limitación, cáncer de pulmón, cáncer de mama, cáncer de próstata, melanoma, carcinoma de células escamosas, cáncer colorrectal, cáncer de páncreas, cáncer de tiroides, cáncer de endometrio, cáncer de vejiga, cáncer de riñón, tumor sólido, leucemia, linfoma no Hodgkin. En una realización, puede incluirse una célula cancerosa resistente a fármacos en el modelo de cáncer. Los ejemplos específicos de líneas celulares que pueden incluirse en un modelo de cáncer incluyen, sin limitación, células PaCa2, HepG2, PC3 y MCF7. Se conocen numerosas líneas celulares de cáncer en la técnica, y cualquiera de tales líneas celulares de cáncer puede incluirse en un modelo de cáncer de la invención.

Los modelos celulares de la invención pueden incluir una o más "células de control". En una realización, una célula de control puede ser una célula cancerosa no tratada o no perturbada. En otra realización, una "célula de control" puede ser una célula normal, no cancerosa. Una cualquiera de las numerosas líneas celulares normales no cancerosas puede incluirse en el modelo celular. En una realización, las células normales son una o más de las células THLE2 y HDFa. En una realización, se incluyen al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 o más tipos de células normales diferentes en el modelo de cáncer. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 líneas o tipos de células normales diferentes. En una realización, las células normales son todas del mismo tipo, por ejemplo, todas células epiteliales o de mama sanas, pero son diferentes líneas celulares establecidas, por ejemplo, diferentes líneas celulares establecidas de células epiteliales o de mama. En una realización, como control, se cultivan una o más líneas celulares no cancerosas normales (por ejemplo, THLE2 y HDFa) en condiciones similares, y/o se exponen a la misma perturbación, como las células cancerosas del modelo celular para identificar proteínas o rutas únicas del cáncer.

Un modelo de cáncer personalizado también puede comprender condiciones de cultivo celular que imitan un aspecto característico del estado o proceso canceroso. Por ejemplo, pueden seleccionarse condiciones de cultivo celular que se aproximen estrechamente a las condiciones de una célula cancerosa en un entorno tumoral *in vivo*, o para imitar diversos aspectos del crecimiento del cáncer, aislando diferentes condiciones de crecimiento de las células cancerosas. En algunos casos, las condiciones de cultivo celular son condiciones de estrés.

Uno de esos "entornos" de cáncer, o condición de estrés, es la hipoxia, una condición que se encuentra normalmente dentro de un tumor sólido. Puede inducirse hipoxia en células en células usando métodos reconocidos en la técnica. Por ejemplo, puede inducirse hipoxia colocando sistemas celulares en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que puede inundarse con una mezcla de gases industriales que contiene el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Pueden medirse los efectos después de un periodo predeterminado, por ejemplo, a las 24 horas después del tratamiento de hipoxia, con y sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM).

Del mismo modo, el tratamiento con ácido láctico de las células imita un entorno celular en el que la actividad de glicólisis es alta, como existe en el entorno tumoral *in vivo*. El estrés inducido por ácido láctico puede investigarse a una concentración de ácido láctico final de aproximadamente 12,5 mM en un tiempo predeterminado, por ejemplo, a las 24 horas, con o sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM).

La hiperglucemia es normalmente una condición que se encuentra en la diabetes; sin embargo, la hiperglucemia también imita en cierta medida un aspecto del crecimiento del cáncer porque muchas células cancerosas dependen de la glucosa como su principal fuente de energía. La exposición de las células objeto a una condición hiperglucémica típica puede incluir la adición del 10% de glucosa de grado de cultivo a medios adecuados, de modo que la concentración final de glucosa en los medios sea de aproximadamente 22 mM.

Pueden investigarse condiciones individuales que reflejan diferentes aspectos del crecimiento del cáncer por separado en el modelo de cáncer personalizado y/o pueden combinarse entre sí. En una realización, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más condiciones que reflejan o simulan diferentes aspectos del crecimiento/estados del cáncer se investigan en el modelo de cáncer personalizado. En una realización, se investigan condiciones individuales y, además, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más de las condiciones que reflejan o simulan diferentes aspectos del crecimiento/estados del cáncer en el modelo de cáncer personalizado. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 ó 10 y 50 condiciones diferentes.

Una vez que se construye el modelo celular personalizado, pueden aplicarse una o más "perturbaciones" al sistema, tales como variación genética de un paciente a otro, o con/sin tratamiento con determinados fármacos o profármacos. Véase la figura 15D. Los efectos de tales perturbaciones en el sistema, incluyendo el efecto sobre las células cancerosas relacionadas con la enfermedad y las células de control normales relacionadas con la enfermedad, pueden medirse usando diversos medios reconocidos en la técnica o patentados, tal como se describe en la sección III.B a continuación.

En un experimento a modo de ejemplo, las líneas de cáncer PaCa2, HepG2, PC3 y MCF7, y las líneas celulares

normales THLE2 y HDFa, se acondicionan en cada una de condiciones de hiperglucemia, hipoxia y ricas en ácido láctico, así como en todas las combinaciones de dos o tres de estas condiciones, y además con o sin perturbación ambiental, específicamente tratamiento con coenzima Q10. A continuación se enumeran las combinaciones a modo de ejemplo de condiciones, con o sin perturbación, tratamiento con coenzima Q10, que pueden usarse para tratar las células cancerosas y/o células de control (por ejemplo, células normales) del modelo celular de cáncer. Pueden formularse fácilmente otras combinaciones dependiendo de la evaluación biológica por interrogación específica que esté realizándose.

5

10

15

20

25

30

35

40

45

50

55

1. Sólo medios
2. CTL coenzima Q10 50 μ M
3. CTL coenzima Q10 100 μ M
4. Ácido láctico 12,5 mM
5. Ácido láctico 12,5 mM + CTL coenzima Q10 50 μ M
6. Ácido láctico 12,5 mM + CTL coenzima Q10 100 μ M
7. Hipoxia
8. Hipoxia + CTL coenzima Q10 50 μ M
9. Hipoxia + CTL coenzima Q10 100 μ M
10. Hipoxia + ácido láctico 12,5 mM
11. Hipoxia + ácido láctico 12,5 mM + CTL coenzima Q10 50 μ M
12. Hipoxia + ácido láctico 12,5 mM + CTL coenzima Q10 100 μ M
13. Medios + glucosa 22 mM
14. CTL coenzima Q10 50 μ M + glucosa 22 mM
15. CTL coenzima Q10 100 μ M + glucosa 22 mM
16. Ácido láctico 12,5 mM + glucosa 22 mM
17. Ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 50 μ M
18. Ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 100 μ M
19. Hipoxia + glucosa 22 mM
20. Hipoxia + glucosa 22 mM + CTL coenzima Q10 50 μ M
21. Hipoxia + glucosa 22 mM + CTL coenzima Q10 100 μ M
22. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM
23. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 50 μ M
24. Hipoxia + ácido láctico 12,5 mM + glucosa 22 mM + CTL coenzima Q10 100 μ M

60

65

En determinadas situaciones, pueden realizarse experimentos de interferencia o SEC entre diferentes células cancerosas (por ejemplo, HepG2 y PaCa2) con diversos propósitos interrelacionados. En algunas realizaciones que implican interferencia, los experimentos realizados en los modelos celulares están diseñados para determinar la modulación del estado celular o la función de un sistema o una población celular (por ejemplo, célula de hepatocarcinoma HepG2) por otro sistema o población celular (por ejemplo, cáncer de páncreas PaCa2) en condiciones de tratamiento definidas (por ejemplo, hiperglucemia, hipoxia (isquemia)). Según una configuración típica, un primer sistema/población celular se pone en contacto con componentes de estímulo externo, tales como una molécula candidata (por ejemplo, una molécula de fármaco pequeña, una proteína) o una condición candidata (por ejemplo, hipoxia, entorno con alto contenido de glucosa). En respuesta, el primer sistema/población celular

cambia su transcriptoma, proteoma, metaboloma y/o interactoma, lo que conduce a cambios que pueden detectarse fácilmente tanto en el interior como en el exterior de la célula. Por ejemplo, los cambios en el transcriptoma pueden medirse mediante el nivel de transcripción de una pluralidad de ARNm diana; los cambios en el proteoma pueden medirse mediante el nivel de expresión de una pluralidad de proteínas diana; y los cambios en el metaboloma pueden medirse mediante el nivel de una pluralidad de metabolitos diana mediante ensayos diseñados específicamente para metabolitos dados. Alternativamente, los cambios a los que se hizo referencia anteriormente en el metaboloma y/o proteoma, al menos con respecto a determinados metabolitos o proteínas secretados, también pueden medirse mediante sus efectos sobre el segundo sistema/población celular, incluyendo la modulación del transcriptoma, proteoma, metaboloma e interactoma del segundo sistema/población celular. Por tanto, los experimentos pueden usarse para identificar los efectos de la(s) molécula(s) de interés secretada(s) por el primer sistema/población celular sobre un segundo sistema/población celular en diferentes condiciones de tratamiento. Los experimentos también pueden usarse para identificar cualquier proteína que se module como resultado de la señalización del primer sistema celular (en respuesta al tratamiento con componente de estímulo externo) a otro sistema celular, por ejemplo, mediante examen diferencial de proteómica. La misma configuración experimental también puede adaptarse para una configuración inversa, de modo que también puedan evaluarse los efectos recíprocos entre los dos sistemas celulares. En general, para este tipo de experimento, la elección de las parejas de líneas celulares se basa en gran medida en factores tales como el origen, estado de la enfermedad y la función celular.

Aunque los sistemas de dos células están implicados normalmente en este tipo de configuración experimental, también pueden diseñarse experimentos similares para más de dos sistemas celulares, por ejemplo, inmovilizando cada sistema celular distinto sobre un soporte sólido independiente.

El modelo de cáncer personalizado puede establecerse y usarse en la totalidad de las etapas de la tecnología de plataforma de la invención para identificar, en última instancia, una relación causal única en el sistema biológico, mediante la realización de las etapas descritas en el presente documento. Sin embargo, el experto en la técnica entenderá que un modelo de cáncer personalizado que se usa para generar una red de relaciones causales de consenso inicial de "primera generación" puede evolucionar o expandirse de manera continua a lo largo del tiempo, por ejemplo, mediante la introducción de líneas celulares de cáncer o normales adicionales y/o estados de cáncer adicionales. Pueden recopilarse datos adicionales del modelo de cáncer evolucionado, es decir, datos de la(s) parte(s) recién añadida(s) del modelo de cáncer. Los nuevos datos recopilados de un modelo de cáncer expandido o evolucionado, es decir, de la(s) parte(s) recién añadida(s) del modelo de cáncer, pueden introducirse luego en los grupos de datos usados previamente para generar la red de relaciones causales de consenso de "primera generación" para generar una red de relaciones causales de consenso de "segunda generación" más robusta. Pueden identificarse nuevas relaciones causales únicas para el estado de cáncer (o únicas para la respuesta del estado de cáncer a una perturbación) a partir de la red de relaciones causales de consenso de "segunda generación". De esta manera, la evolución del modelo de cáncer proporciona una evolución de las redes de relaciones causales de consenso, lo que proporciona información nueva y/o más fiable sobre los controladores determinantes (o moduladores) del estado del cáncer.

(iii) Modelos celulares de diabetes/obesidad/enfermedad cardiovascular

Otros ejemplos de un sistema o proceso biológico son la diabetes, la obesidad y la enfermedad cardiovascular. Al igual que con el cáncer, los estados patológicos relacionados de diabetes, obesidad y enfermedad cardiovascular son estados patológicos complicados caracterizados por múltiples aspectos únicos. Sería de interés identificar las proteínas/rutas que controlan la patogenia de la diabetes/obesidad/enfermedad cardiovascular. También sería interesante descifrar la respuesta única de las células asociadas a diabetes/obesidad/enfermedad cardiovascular al tratamiento farmacológico en comparación con las respuestas de las células normales. Con este propósito, puede establecerse un modelo personalizado de diabetes/obesidad/cardiovascular para simular un entorno experimentado por las células relevantes para la enfermedad, eligiendo líneas celulares apropiadas y creando condiciones de cultivo celular que imiten un aspecto característico del estado o proceso de enfermedad. Por ejemplo, pueden seleccionarse condiciones de cultivo celular que se aproximen estrechamente a la hiperglucemia, hiperlipidemia, hiperinsulinemia, hipoxia o condiciones ricas en ácido láctico.

Cualquier célula relevante para la diabetes/obesidad/enfermedad cardiovascular puede incluirse en el modelo de diabetes/obesidad/enfermedad cardiovascular. Los ejemplos de células relevantes para la diabetes/obesidad/enfermedad cardiovascular incluyen, por ejemplo, adipocitos, miotubos, hepatocitos, células de músculo liso aórtico (HASMC) y células tubulares proximales (por ejemplo, HK2). Pueden incluirse múltiples tipos de células del mismo o diferente origen, en oposición a un único tipo de células, en el modelo de diabetes/obesidad/enfermedad cardiovascular. En una realización, se incluyen al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 o más tipos de células diferentes en el modelo de diabetes/obesidad/enfermedad cardiovascular. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 tipos de células diferentes. En una realización, las células son todas del mismo tipo, por ejemplo, todas adipocitos, pero son diferentes líneas celulares establecidas, por ejemplo, diferentes líneas celulares de adipocitos establecidas. Muchos otros tipos de células que están implicados en el estado de la

diabetes/obesidad/enfermedad cardiovascular se conocen en la técnica, y cualquiera de tales células puede incluirse en un modelo de diabetes/obesidad/enfermedad cardiovascular de la invención.

Los modelos celulares de diabetes/obesidad/enfermedad cardiovascular de la invención pueden incluir una o más "células de control". En una realización, una célula de control puede ser una célula relevante para la enfermedad no tratada o no perturbada, por ejemplo, una célula que no está expuesta a una condición hiperlipidémica o hiperinsulinémica. En otra realización, una "célula de control" puede ser una célula no relevante para la enfermedad, tal como una célula epitelial. Una cualquiera de las numerosas células no relevantes para la enfermedad puede incluirse en el modelo celular. En una realización, se incluyen al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 o más tipos de células diferentes no relevantes para la enfermedad en el modelo celular. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, o 5 y 15 diferentes líneas celulares o tipos de células no relevantes para la enfermedad diferentes. En una realización, las células no relevantes para la enfermedad son todas del mismo tipo, por ejemplo, todas células epiteliales o de mama sanas, pero son diferentes líneas celulares establecidas, por ejemplo, diferentes líneas celulares establecidas de células epiteliales o de mama. En una realización, como control, una o más líneas celulares no relevantes para la enfermedad se cultivan en condiciones similares, y/o se exponen a la misma perturbación que las células relevantes para la enfermedad del modelo celular para identificar proteínas o rutas únicas para diabetes/obesidad/enfermedad cardiovascular.

Un modelo personalizado de diabetes/obesidad/enfermedad cardiovascular también puede comprender condiciones de cultivo celular que imitan un aspecto característico de (representan la fisiopatología de) el estado o proceso de enfermedad de diabetes/obesidad/enfermedad cardiovascular. Por ejemplo, pueden seleccionarse condiciones de cultivo celular que se aproximen estrechamente a las condiciones de una célula relevante para diabetes/obesidad/enfermedad cardiovascular en su entorno *in vivo*, o para imitar diversos aspectos de la diabetes/obesidad/enfermedad cardiovascular. En algunos casos, las condiciones de cultivo celular son condiciones de estrés.

Las condiciones a modo de ejemplo que representan la fisiopatología de la diabetes/obesidad/enfermedad cardiovascular incluyen, por ejemplo, una o más de hipoxia, condiciones ricas en ácido láctico, hiperglucemia, hiperlipidemia e hiperinsulinemia. Puede inducirse hipoxia en células en células usando métodos reconocidos en la técnica. Por ejemplo, puede inducirse hipoxia colocando sistemas celulares en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que puede inundarse con una mezcla de gases industriales que contiene el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Pueden medirse los efectos después de un periodo predeterminado, por ejemplo, a las 24 horas después del tratamiento de hipoxia, con y sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM).

Asimismo, el tratamiento con ácido láctico de las células imita un entorno celular en el que la actividad de glicólisis es alta. El estrés inducido por ácido láctico puede investigarse a una concentración de ácido láctico final de aproximadamente 12,5 mM en un tiempo predeterminado, por ejemplo, a las 24 horas, con o sin componentes de estímulo externo adicionales (por ejemplo, CoQ10 a 0, 50 ó 100 µM). La hiperglucemia es una condición que se encuentra en la diabetes. La exposición de las células objeto a una condición hiperglucémica típica puede incluir la adición del 10% de glucosa de grado de cultivo a los medios adecuados, de modo que la concentración final de glucosa en los medios sea de aproximadamente 22 mM. La hiperlipidemia es una condición que se encuentra en la obesidad y la enfermedad cardiovascular. Las condiciones hiperlipidémicas pueden proporcionarse cultivando células en medios que contienen palmitato de sodio 0,15 mM. La hiperinsulinemia es una condición que se encuentra en la diabetes. Las condiciones hiperinsulinémicas pueden inducirse mediante el cultivo de las células en medios que contienen insulina 1000 nM.

Las condiciones adicionales que representan la fisiopatología de la diabetes/obesidad/enfermedad cardiovascular incluyen, por ejemplo, una o más de inflamación, estrés del retículo endoplasmático, estrés mitocondrial y estrés peroxisómico. Los métodos para crear una condición de tipo inflamatoria en las células se conocen en la técnica. Por ejemplo, una condición inflamatoria puede simularse cultivando células en presencia de TNFalfa y/o IL-6. Los métodos para crear condiciones que simulan el estrés del retículo endoplasmático también se conocen en la técnica. Por ejemplo, pueden crearse condiciones que simulan el estrés del retículo endoplasmático mediante el cultivo de células en presencia de tapsigargina y/o tunicamicina. Los métodos para crear condiciones que simulan el estrés mitocondrial también se conocen en la técnica. Por ejemplo, pueden crearse condiciones que simulan el estrés mitocondrial cultivando células en presencia de rapamicina y/o galactosa. Los métodos para crear condiciones que simulan el estrés peroxisómico también se conocen en la técnica. Por ejemplo, pueden crearse condiciones que simulan el estrés peroxisómico cultivando células en presencia de ácido abscísico.

Las condiciones individuales que reflejan diferentes aspectos de la diabetes/obesidad/enfermedad cardiovascular pueden investigarse por separado en el modelo personalizado de diabetes/obesidad/enfermedad cardiovascular, y/o pueden combinarse entre sí. En una realización, se investigan combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más condiciones que reflejan o simulan diferentes aspectos de diabetes/obesidad/enfermedad cardiovascular en el modelo personalizado de diabetes/obesidad/enfermedad

cardiovascular. En una realización, se investigan condiciones individuales y, además, combinaciones de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 o más de las condiciones que reflejan o simulan diferentes aspectos de la diabetes/obesidad/enfermedad cardiovascular en el modelo personalizado de diabetes/obesidad/enfermedad cardiovascular. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 o 10 y 50 condiciones diferentes.

Una vez que se construye el modelo celular personalizado, pueden aplicarse una o más "perturbaciones" al sistema, tales como variación genética de un paciente a otro, o con/sin tratamiento con determinados fármacos o profármacos. Véase la figura 15D. Los efectos de tales perturbaciones en el sistema, incluyendo el efecto sobre las células relacionadas con la diabetes/obesidad/enfermedad cardiovascular, pueden medirse usando diversos medios reconocidos en la técnica o patentados, tal como se describe en la sección III.B a continuación.

En un experimento a modo de ejemplo, cada uno de adipocitos, miotubos, hepatocitos, células de músculo liso aórtico (HASMC) y células tubulares proximales (HK2), se acondicionan en cada una de las condiciones de hiperglucemia, hipoxia, hiperlipidemia, hiperinsulinemia y ricas en ácido láctico, así como en todas las combinaciones de dos, tres, cuatro y las cinco condiciones, y además con o sin perturbación ambiental, específicamente el tratamiento con coenzima Q10. Además de las combinaciones a modo de ejemplo de condiciones descritas anteriormente en el contexto del modelo de cáncer, a continuación se enumeran algunas combinaciones a modo de ejemplo adicionales de condiciones, con o sin una perturbación, por ejemplo, tratamiento con coenzima Q10, que pueden usarse para tratar las células relevantes para la enfermedad de diabetes/obesidad/enfermedad cardiovascular (y/o células de control) del modelo celular de diabetes/obesidad/enfermedad cardiovascular. Estas están destinadas simplemente a ser a modo de ejemplo, y el experto en la técnica apreciará que cualquier condición individual y/o combinación de las condiciones mencionadas anteriormente que represente la fisiopatología de la diabetes/obesidad/enfermedad cardiovascular puede emplearse en el modelo celular para producir grupos de datos de salida. Pueden formularse fácilmente otras combinaciones dependiendo de la evaluación biológica por interrogación específica que esté realizándose.

1. Sólo medios

2. CTL Coenzima Q10 50 μ M

3. CTL Coenzima Q10 100 μ M

4. Palmitato de sodio 0,15 mM

5. Palmitato de sodio 0,15 mM + coenzima CTL Q10 50 μ M

6. Palmitato de sodio 0,15 mM + coenzima CTL Q10 100 μ M

7. Insulina 1000 nM

8. Insulina 1000 nM + coenzima CTL Q10 50 μ M

9. Insulina 1000 nM + coenzima CTL Q10 100 μ M

10. Insulina 1000 nM + palmitato de sodio 0,15 mM

11. Insulina 1000 nM + palmitato de sodio 0,15 mM + coenzima CTL Q10 50 μ M

12. Insulina 1000 nM + palmitato de sodio 0,15 mM + coenzima CTL Q10 100 μ M

En determinadas situaciones, pueden realizarse experimentos de interferencia o SEC entre diferentes células relevantes para la enfermedad (por ejemplo, células HASMC y HK2, o células hepáticas y adipocitos) con diversos propósitos interrelacionados. En algunas realizaciones que implican interferencia, los experimentos realizados en los modelos celulares están diseñados para determinar la modulación del estado celular o la función de un sistema o una población celular (por ejemplo, células hepáticas) por otro sistema o población celular (por ejemplo, adipocitos) en condiciones de tratamiento definidas (por ejemplo, hiperglucemia, hipoxia, hiperlipidemia, hiperinsulinemia). Según una configuración típica, un primer sistema/población celular se pone en contacto con componentes de estímulo externo, tales como una molécula candidata (por ejemplo, una molécula de fármaco pequeña, una proteína) o una condición candidata (por ejemplo, hipoxia, entorno con alto contenido de glucosa). En respuesta, el primer sistema/población celular cambia su transcriptoma, proteoma, metaboloma y/o interactoma, lo que conduce a cambios que pueden detectarse fácilmente tanto en el interior como en el exterior de la célula. Por ejemplo, los cambios en el transcriptoma pueden medirse mediante el nivel de transcripción de una pluralidad de ARNm diana; los cambios en el proteoma pueden medirse mediante el nivel de expresión de una pluralidad de proteínas diana; y

los cambios en el metaboloma pueden medirse mediante el nivel de una pluralidad de metabolitos diana mediante ensayos diseñados específicamente para metabolitos dados. Alternativamente, los cambios a los que se hizo referencia anteriormente en el metaboloma y/o proteoma, al menos con respecto a determinados metabolitos o proteínas secretados, también pueden medirse mediante sus efectos sobre el segundo sistema/población celular, incluyendo la modulación del transcriptoma, proteoma, metaboloma e interactoma del segundo sistema/población celular. Por tanto, los experimentos pueden usarse para identificar los efectos de las molécula(s) de interés secretada(s) por el primer sistema/población celular sobre un segundo sistema/población celular en diferentes condiciones de tratamiento. Los experimentos también pueden usarse para identificar cualquier proteína que se module como resultado de la señalización del primer sistema celular (en respuesta al tratamiento con componente de estímulo externo) a otro sistema celular, por ejemplo, mediante examen diferencial de proteómica. La misma configuración experimental también puede adaptarse para una configuración inversa, de modo que también puedan evaluarse los efectos recíprocos entre los dos sistemas celulares. En general, para este tipo de experimento, la elección de las parejas de líneas celulares se basa en gran medida en factores tales como el origen, el estado de la enfermedad y la función celular.

Aunque los sistemas de dos células están implicados normalmente en este tipo de configuración experimental, también pueden diseñarse experimentos similares para más de dos sistemas celulares, por ejemplo, inmovilizando cada sistema celular distinto sobre un soporte sólido independiente.

El modelo personalizado de diabetes/obesidad/enfermedad cardiovascular puede establecerse y usarse en la totalidad de las etapas de la tecnología de plataforma de la invención para identificar, en última instancia, una relación causal única para el estado de la diabetes/obesidad/enfermedad cardiovascular, llevando a cabo las etapas descritas en el presente documento. Sin embargo, el experto en la técnica entenderá que, al igual que con un modelo de cáncer, un modelo personalizado de diabetes/obesidad/enfermedad cardiovascular que se usa para generar una red de relaciones causales de consenso inicial de "primera generación" puede evolucionar o expandirse de manera continua a lo largo del tiempo, por ejemplo, mediante la introducción de líneas celulares relevantes para la enfermedad adicionales y/o condiciones relevantes para la enfermedad adicionales. Pueden recopilarse datos adicionales del modelo evolucionado de diabetes/obesidad/enfermedad cardiovascular, es decir, datos de la(s) parte(s) recién añadida(s) del modelo de cáncer. Los nuevos datos recopilados de un modelo expandido o evolucionado, es decir, de la(s) parte(s) recién añadida(s) del modelo, pueden introducirse luego en los grupos de datos usados previamente para generar la red de relaciones causales de consenso de "primera generación" para generar un red de relaciones causales de consenso de "segunda generación" más robusta. Pueden identificarse nuevas relaciones causales únicas para el estado de diabetes/obesidad/enfermedad cardiovascular (o únicas para la respuesta del estado de diabetes/obesidad/enfermedad cardiovascular a una perturbación) a partir de la red de relaciones causales de consenso de "segunda generación". De esta manera, la evolución del modelo de diabetes/obesidad/enfermedad cardiovascular proporciona una evolución de las redes de relaciones causales de consenso, proporcionando así información nueva y/o más fiable sobre los controladores determinantes (o moduladores) del estado de la diabetes/obesidad/enfermedad cardiovascular.

B. Uso de modelos celulares para evaluaciones biológicas por interrogación

Los métodos y modelos celulares proporcionados en la presente invención pueden usarse para, o aplicarse a, cualquier número de "evaluaciones biológicas por interrogación". El uso de los métodos de la invención para una evaluación biológica por interrogación facilita la identificación de "moduladores" o "controladores" de proceso celular determinantes de un sistema biológico.

Tal como se usa en el presente documento, una "evaluación biológica por interrogación" puede incluir la identificación de uno o más moduladores de un sistema biológico, por ejemplo, "controladores" de proceso celular determinantes (por ejemplo, un aumento o disminución de la actividad de una ruta biológica, o miembros clave de la ruta, o reguladores clave para miembros de la ruta) asociados con la perturbación ambiental o el componente de estímulo externo, o una única relación causal única en un sistema o proceso biológico. Puede incluir además etapas adicionales diseñadas para someter a prueba o verificar si los controladores de proceso celular determinantes identificados son necesarios y/o suficientes para los eventos posteriores asociados con la perturbación ambiental o el componente de estímulo externo, incluyendo modelos animales *in vivo* y/o experimentos de cultivo tisular *in vitro*.

En determinadas realizaciones, la evaluación biológica por interrogación es el diagnóstico o la estadificación de un estado de enfermedad, en la que los moduladores identificados de un sistema biológico, por ejemplo, controladores de proceso celular determinantes (por ejemplo, diferenciales de interferencia o relaciones causales únicas en un sistema o proceso biológico)) representan o bien marcadores de enfermedad o bien dianas terapéuticas que pueden ser objeto de intervención terapéutica. La evaluación biológica por interrogación objeto es adecuada para cualquier estado de enfermedad en teoría, pero puede resultar particularmente útil en áreas como oncología/biología del cáncer, diabetes, obesidad, enfermedad cardiovascular y estados neurológicos (especialmente enfermedades neurodegenerativas, tales como, sin limitación, Enfermedad de Alzheimer, enfermedad de Parkinson, enfermedad de Huntington, esclerosis lateral amiotrófica (ELA) y neurodegeneración relacionada con el envejecimiento).

En determinadas realizaciones, la evaluación biológica por interrogación es la determinación de la eficacia de un

fármaco, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia o relaciones causales únicas en un sistema o proceso biológico)) pueden ser el sello distintivo de un fármaco exitoso y pueden usarse, a su vez, para identificar agentes adicionales, tales como MIM o intercambiadores epimetabólicos, para tratar la misma enfermedad.

5 En determinadas realizaciones, la evaluación biológica por interrogación es la identificación de dianas farmacológicas para prevenir o tratar la infección (por ejemplo, infección bacteriana o viral), en el que el controlador de proceso celular determinante identificado (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores/indicadores o moléculas biológicas clave causantes del estado infeccioso y pueden usarse, a su vez, para identificar agentes antiinfecciosos.

10 En determinadas realizaciones, la evaluación biológica por interrogación es la evaluación de un efecto molecular de un agente, por ejemplo, un fármaco, sobre un perfil de enfermedad dado, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia o las relaciones causales únicas en un sistema o proceso biológico) pueden ser un aumento o disminución de la actividad de una o más rutas biológicas, o miembros clave de la(s) ruta(s), o reguladores clave para los miembros de la(s) ruta(s) y pueden usarse, a su vez, por ejemplo para predecir la eficacia terapéutica del agente para la enfermedad dada.

15 En determinadas realizaciones, la evaluación biológica por interrogación es la evaluación del perfil toxicológico de un agente, por ejemplo, un fármaco, sobre una célula, un tejido, órgano u organismo, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser indicadores de toxicidad, por ejemplo, citotoxicidad y pueden usarse, a su vez, para predecir o identificar el perfil toxicológico del agente. En una realización, los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) es un indicador de cardiotoxicidad de un fármaco o candidato a fármaco, y pueden usarse, a su vez, para predecir o identificar el perfil cardiotoxicológico del fármaco o candidato a fármaco.

20 En determinadas realizaciones, la evaluación biológica por interrogación es la identificación de dianas farmacológicas para prevenir o tratar una enfermedad o un trastorno provocado por armas biológicas, tales como protozoos, hongos, bacterias, protistas, virus o toxinas que provocan enfermedad, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores/indicadores o moléculas biológicas clave causantes de dicha enfermedad o dicho trastorno y pueden usarse, a su vez, para identificar agentes de biodefensa.

25 En determinadas realizaciones, la evaluación biológica por interrogación es la identificación de dianas para agentes antienvjecimiento, tales como cosméticos antienvjecimiento, en la que los moduladores identificados de un sistema biológico, por ejemplo, controlador de proceso celular determinante (por ejemplo, diferenciales de interferencia celular o relaciones causales únicas en un sistema o proceso biológico) pueden ser marcadores o indicadores del proceso de envejecimiento, particularmente el proceso de envejecimiento en la piel y pueden usarse, a su vez, para identificar agentes antienvjecimiento.

30 En un modelo celular a modo de ejemplo para el envejecimiento que se usa en los métodos de la invención para identificar dianas para cosméticos antienvjecimiento, el modelo celular comprende una célula epitelial envejecida que, por ejemplo, se trata con luz UV (una perturbación ambiental o un componente de estímulo externo), y/o células neonatales, que también se tratan opcionalmente con luz UV. En una realización, un modelo celular para el envejecimiento comprende un sistema de interferencia celular. En un sistema de interferencia de dos células a modo de ejemplo establecido para identificar dianas para cosméticos antienvjecimiento, una célula epitelial envejecida (primer sistema celular) puede tratarse con luz UV (un componente de estímulo externo) y pueden medirse los cambios, por ejemplo, cambios proteómicos y/o cambios funcionales, en una célula neonatal (segundo sistema celular) resultantes del contacto de las células neonatales con el medio condicionado de la célula epitelial envejecida tratada, por ejemplo, pueden medirse los cambios de proteoma usando espectrometría de masas cuantitativa convencional, o puede identificarse una relación causal única en el envejecimiento a partir de una red de relaciones causales generada a partir de los datos.

V. Análisis de muestras proteómicas

35 En determinadas realizaciones, el método objeto emplea análisis proteómico cuantitativo de alto rendimiento a gran escala de cientos de muestras de carácter similar, y proporciona los datos necesarios para identificar los diferenciales de salida celular.

40 Existen numerosas tecnologías reconocidas en la técnica adecuadas con este propósito. A continuación se describe en resumen una técnica a modo de ejemplo, el análisis de iTRAQ en combinación con la espectrometría de masas.

Para proporcionar muestras de referencia para la cuantificación relativa con la técnica de iTRAQ, se crean múltiples reservas de QC. Se generaron dos reservas de QC independientes, que consistían en alícuotas de cada muestra, a partir de las muestras de célula # 1 y célula # 2; estas muestras se designan como QCS1 y QCS2, y QCP1 y QCP2 para sobrenadantes y sedimentos, respectivamente. Con el propósito de permitir la comparación de la concentración de proteína a través de las dos líneas celulares, se combinan alícuotas de sedimento celular de las reservas de QC descritas anteriormente en volúmenes iguales para generar muestras de referencia (QCP).

El enfoque de la proteómica cuantitativa se basa en el etiquetado de isótopos estables con el reactivo de iTRAQ 8-plex y 2D-CL MALDI EM/EM para la identificación y cuantificación de péptidos. La cuantificación con esta técnica es relativa: a los péptidos y proteínas se les asignan proporciones de abundancia en relación con una muestra de referencia. Las muestras de referencia comunes en múltiples experimentos de iTRAQ facilitan la comparación de muestras en múltiples experimentos de iTRAQ.

Para implementar este esquema de análisis, se combinan seis muestras primarias y dos muestras de reserva de control en una mezcla de iTRAQ 8-plex, marcándose las muestras de reserva de control con 113 y 117 reactivos según las sugerencias del fabricante. Esta mezcla de ocho muestras se fracciona luego mediante cromatografía de líquidos bidimensional; intercambio catiónico fuerte (SCX) en la primera dimensión y HPLC de fase inversa en la segunda dimensión. El eluyente de HPLC se fracciona directamente sobre placas para MALDI, y las placas se analizan en un espectrómetro de masas MDS SCIEX/AB 4800 MALDI TOF/TOF.

En ausencia de información adicional, se supone que los cambios más importantes en la expresión de proteína son aquellos dentro de los mismos tipos de células en diferentes condiciones de tratamiento. Por este motivo, las muestras primarias de la célula # 1 y la célula # 2 se analizan en mezclas de iTRAQ independientes. Para facilitar la comparación de la expresión de proteína en las muestras de la célula # 1 frente a la célula # 2, se analizan muestras universales de QCP en las "ranuras de iTRAQ" disponibles que no están ocupadas por las muestras de QC primarias o específicas de líneas celulares (QC1 y QC2).

En el presente documento se proporciona una breve descripción de los procedimientos de laboratorio empleados.

A. Extracción de proteína de muestras de sobrenadante celular

Para las muestras de sobrenadante celular (CSN), las proteínas del medio de cultivo están presentes en un gran exceso con respecto a las proteínas secretadas por las células cultivadas. En un intento por reducir este fondo, se implementó la reducción inicial de proteína abundante. Como no están disponibles columnas de afinidad específicas para proteínas séricas bovinas o equinas, se usó una columna de IgY14 anti-humano. Aunque los anticuerpos están dirigidos contra proteínas humanas, se anticipó que la amplia especificidad proporcionada por la naturaleza policlonal de los anticuerpos logra la reducción de las proteínas tanto bovinas como equinas presentes en los medios de cultivo celular que se usaron.

Se carga una alícuota de 200 µl del material de CSN QC en una columna de reducción de IgY14 de 10 ml antes del inicio del estudio para determinar la concentración de proteína total (ensayo de ácido bicinónico (BCA)) en el material de flujo directo. Luego se selecciona el volumen de carga para lograr una fracción reducida que contiene aproximadamente 40 µg de proteína total.

B. Extracción de proteína de sedimentos celulares

Se lisa una alícuota de la célula # 1 y la célula # 2 en el tampón de lisis "convencional" usado para el análisis de muestras de tejido en BGM, y se determina el contenido de proteína total mediante el ensayo de BCA. Una vez establecido el contenido de proteína de estos lisados celulares representativos, se procesaron todas las muestras de sedimentos celulares (incluyendo las muestras de QC descritas en la sección 1.1) para dar lisados celulares. Se transfirieron cantidades de lisado de aproximadamente 40 µg de proteína total al flujo de trabajo de procesamiento.

C. Preparación de muestras para espectrometría de masas

La preparación de muestras sigue procedimientos normalizados de trabajo y constituye lo siguiente:

- Reducción y alquilación de proteínas
- Limpieza de proteínas en columna de fase inversa (sólo sedimentos celulares)
- Digestión con tripsina
- Etiquetado con iTRAQ
- Cromatografía de intercambio catiónico fuerte: colección de seis fracciones (sistema Agilent 1200)

- Fraccionamiento mediante HPLC y aplicación en placas para MALDI (sistema Dionex Ultimate3000/Probot)

D. MALDI EM y EM/EM

5 HPLC-EM emplea generalmente estrategias de ESI EM/EM en línea. BG Medicine usa una plataforma de CL/MALDI EM/EM fuera de línea que da como resultado una mejor concordancia de los grupos de proteínas observados a través de las muestras primarias sin la necesidad de inyectar la misma muestra varias veces. Después de la recopilación de datos de primer pase en todas las mezclas de iTRAQ, dado que las fracciones de péptidos se retienen en las placas objetivo para MALDI, las muestras pueden analizarse por segunda vez usando un patrón de adquisición de EM/EM dirigido derivado del conocimiento obtenido durante la primera adquisición. De esta manera, se logra la frecuencia de observación máxima para todas las proteínas identificadas (de manera ideal, cada proteína debe medirse en cada mezcla de iTRAQ).

15 E. Procesamiento de datos

El proceso de procesamiento de datos dentro del flujo de trabajo de BGM Proteomics puede separarse en aquellos procedimientos, tales como identificación y cuantificación preliminar de péptidos que se completan para cada mezcla de iTRAQ individualmente (sección 1.5.1) y aquellos procesos (sección 1.5.2), tales como asignación final de péptidos a proteínas y cuantificación final de proteínas, que no se completan hasta que se completa la adquisición de datos para el proyecto.

Los principales etapas de procesamiento de datos dentro del flujo de trabajo de BGM Proteomics son:

- Identificación de péptidos usando el motor de búsqueda de la base de datos Mascot (Matrix Sciences)
- Validación automatizada interna de ID de Mascot
- Cuantificación de péptidos y cuantificación preliminar de proteínas
- Tratamiento experto del grupo de datos final
- Asignación final de péptidos de cada mezcla en un grupo común de proteínas usando la herramienta PVT automatizada
- Eliminación de valores aberrantes y cuantificación final de proteínas

(i) Procesamiento de datos de mezclas de iTRAQ individuales

40 A medida que cada mezcla de iTRAQ se procesa a través del flujo de trabajo, los espectros de EM/EM se analizan usando herramientas de software BGM patentadas para la identificación de péptidos y proteínas, así como la evaluación inicial de la información de cuantificación. Basándose en los resultados de este análisis preliminar, se juzga la calidad del flujo de trabajo para cada muestra primaria en la mezcla frente a un grupo de métricas de rendimiento de BGM. Si una muestra (o mezcla) dada no supera las métricas de rendimiento mínimas especificadas, y hay material adicional disponible, esa muestra se repite en su totalidad y son los datos de esta segunda implementación del flujo de trabajo los que se incorporan en el grupo de datos final.

(ii) Identificación de péptidos

50 Se hizo una búsqueda de los espectros de EM/EM en la base de datos de secuencias de proteína Uniprot que contiene secuencias humanas, bovinas y equinas aumentadas por secuencias contaminantes comunes tales como tripsina porcina. Los detalles de los parámetros de búsqueda de Mascot, incluyendo la lista completa de modificaciones, se proporcionan en la Tabla 3.

55 Tabla 3: Parámetros de búsqueda de Mascot

Tolerancia de masa de precursor	100 ppm
Tolerancia de masa de fragmento	0,4 Da
Modificaciones variables	iTRAQ8 de extremo N-terminal
	iTRAQ8 de lisina
	carbamidometilo de Cys

	Piro-Glu (extremo N-terminal)
	Piro-carbamidometilo de Cys (extremo N-terminal)
	Desamidación (sólo N)
	Oxidación (M)
Especificidad de la enzima	Totalmente triptica
Número de sitios tripticos que faltan permitidos	2
Rango de péptido considerado	1

Una vez completada la búsqueda de Mascot, se usa un procedimiento de validación automática para promover (es decir, validar) coincidencias específicas de péptidos de Mascot. La diferenciación entre coincidencias válidas e inválidas se basa en la puntuación de Mascot obtenida en relación con la puntuación de Mascot esperada y la diferencia entre las puntuaciones de Mascot de los péptidos de rango 1 y la de los péptidos de rango 2. Los criterios requeridos para la validación son algo relajados si el péptido es uno de varios coincidentes con una única proteína en la mezcla de iTRAQ o si el péptido está presente en un catálogo de péptidos validados previamente.

(iii) Cuantificación de péptidos y proteínas

El grupo de péptidos validados para cada mezcla se usa para calcular las métricas de cuantificación de proteínas preliminares para cada mezcla. Se calculan proporciones de péptidos dividiendo el área de pico de la etiqueta de iTRAQ (es decir, m/z 114, 115, 116, 118, 119 ó 121) para cada péptido validado entre la mejor representación del área de pico de la reserva de referencia (QC1 o QC2). Esta área de pico es el promedio de los picos de 113 y 117 siempre que ambas muestras pasen los criterios de aceptación de QC. Las proporciones de proteínas preliminares se determinan calculando la mediana de proporción de todos los péptidos validados "útiles" que coinciden con esa proteína. Los péptidos "útiles" están completamente etiquetados con iTRAQ (todos los extremos N-terminales están etiquetados o bien con lisina o bien con piroGlu) y completamente etiquetados con cisteína (es decir, todos los residuos de Cys se alquilan con carbamidometilo o piro-cmc N-terminal).

(iv) Procesamiento posterior a la adquisición

Una vez que se completan todos los pases de adquisición de datos de EM/EM para cada mezcla en el proyecto, los datos se recopilan usando las tres etapas que se analizan a continuación, que tienen como objetivo permitir que los resultados de cada muestra primaria se comparen de manera simple y significativa con los de otra.

(v) Asignación global de secuencias peptídicas a proteínas

La asignación final de las secuencias peptídicas a los números de registro de proteínas se lleva a cabo a través de la herramienta de validación de proteínas (PVT) patentada. El procedimiento de PVT determina el mejor grupo de proteínas mínimo no redundante para describir la colección completa de péptidos identificados en el proyecto. Este es un procedimiento automatizado que se ha optimizado para manejar datos de una taxonomía homogénea.

Las asignaciones de proteína para los experimentos de sobrenadantes se seleccionaron manualmente para tratar las complejidades de las taxonomías mixtas en la base de datos. Dado que el paradigma automatizado no es válido para cultivos celulares hechos crecer en medios complementados con suero bovino y equino, es necesario un tratamiento manual exhaustivo para minimizar la ambigüedad de la fuente de cualquier proteína dada.

(vi) Normalización de las proporciones de péptidos

Las proporciones de péptidos para cada muestra se normalizan según el método de Vandesompele *et al.* Genome Biology, 2002, 3(7), Research 0034.1-11. Este procedimiento se aplica únicamente a las mediciones de sedimentos celulares. Para las muestras de sobrenadante, los datos cuantitativos no se normalizan teniendo en cuenta la mayor contribución a las identificaciones de péptidos procedentes de los medios.

(vii) Cálculo final de las proporciones de proteína

Se usa un procedimiento de eliminación de valores aberrantes estadísticos convencional para eliminar valores aberrantes de alrededor de cada mediana de proporción de proteína, más allá del nivel de $1,96 \sigma$ en el grupo de datos transformados logarítmicamente. Después de este proceso de eliminación, el grupo final de proporciones de proteína se (re)calcula.

VI. Marcadores de la invención y usos de los mismos.

La presente invención se basa, al menos en parte, en la identificación de nuevos biomarcadores que están

asociados con un sistema biológico, tal como un proceso de enfermedad, o la respuesta de un sistema biológico a una perturbación, tal como un agente terapéutico.

En particular, la invención se refiere a marcadores (más adelante en el presente documento “marcadores” o “marcadores de la invención”), que se describen en los ejemplos. La invención proporciona ácidos nucleicos y proteínas que están codificadas por o corresponden a los marcadores (más adelante en el presente documento “ácidos nucleicos marcadores” y “proteínas marcadoras”, respectivamente). Estos marcadores son particularmente útiles en el diagnóstico de estados de enfermedad; pronóstico de estados de enfermedad; desarrollo de dianas farmacológicas para diversos estados de enfermedad; examen para determinar la presencia de toxicidad, preferiblemente toxicidad inducida por fármacos, por ejemplo, cardiotoxicidad; identificación de un agente que provoca o está en riesgo de provocar toxicidad; identificación de un agente que pueda reducir o prevenir la toxicidad inducida por fármacos; alivio, reducción o prevención de la cardiotoxicidad inducida por fármacos; e identificación de marcadores predictivos de cardiotoxicidad inducida por fármacos.

Un “marcador” es un gen cuyo nivel de expresión alterado en un tejido o una célula con respecto a su nivel de expresión en un tejido o una célula normal o sano está asociado a un estado de enfermedad tal como cáncer, diabetes, obesidad, enfermedad cardiovascular o un estado de toxicidad, tal como una toxicidad inducida por fármacos, por ejemplo, cardiotoxicidad. Un “ácido nucleico marcador” es un ácido nucleico (por ejemplo, ARNm, ADNc) codificado por o correspondiente a un marcador de la invención. Tales ácidos nucleicos marcadores incluyen ADN (por ejemplo, ADNc) que comprende la secuencia completa o parcial de cualquiera de los genes que son marcadores de la invención o el complemento de tal secuencia. Tales secuencias las conocen los expertos en la técnica y pueden hallarse, por ejemplo, en el sitio web Pubmed gubernamental de los NIH. Los ácidos nucleicos marcadores también incluyen ARN que comprende la secuencia completa o parcial de cualquiera de los marcadores genéticos de la invención o el complemento de tal secuencia, en el que todos los residuos de timidina se reemplazan por residuos de uridina. Una “proteína marcadora” es una proteína codificada por o correspondiente a un marcador de la invención. Una proteína marcadora comprende la secuencia completa o parcial de cualquiera de las proteínas marcadoras de la invención. Tales secuencias las conocen los expertos en la técnica y pueden hallarse, por ejemplo, en el sitio web Pubmed gubernamental de los NIH. Los términos “proteína” y “polipéptido” se usan indistintamente.

Un líquido corporal de “asociado a estado de enfermedad o estado tóxico” es un líquido que, cuando se encuentra en el cuerpo de un paciente, entra en contacto o pasa a través de células de sarcoma o al que son capaces de pasar células o proteínas que se desprenden de células de sarcoma. Los líquidos corporales asociados a estado de enfermedad o estado tóxico a modo de ejemplo incluyen líquidos sanguíneos (por ejemplo, sangre completa, suero sanguíneo, sangre a la que se han eliminado las plaquetas), y se describen con más detalle a continuación. Los líquidos corporales asociados a estado de enfermedad o estado tóxico no se limitan a sangre completa, sangre a la que se han eliminado las plaquetas, linfa, líquido prostático, orina y semen.

El nivel de expresión “normal” de un marcador es el nivel de expresión del marcador en células de un sujeto o paciente humano no afectado por un estado de enfermedad o un estado de toxicidad.

Una “sobrexpresión” o “mayor nivel de expresión” de un marcador se refiere a un nivel de expresión en una muestra de prueba que es mayor que el error convencionales del ensayo empleado para evaluar la expresión, y es preferiblemente al menos dos veces, y más preferiblemente tres, cuatro, cinco, seis, siete, ocho, nueve o diez veces el nivel de expresión del marcador en una muestra de control (por ejemplo, muestra de un sujeto sano que no tiene el marcador asociado a un estado de enfermedad o un estado de toxicidad, por ejemplo, cáncer, diabetes, obesidad, enfermedad cardiovascular y cardiotoxicidad) y, preferiblemente, el nivel de expresión promedio del marcador en varias muestras de control.

Un “menor nivel de expresión” de un marcador se refiere a un nivel de expresión en una muestra de prueba que es al menos dos veces, y más preferiblemente tres, cuatro, cinco, seis, siete, ocho, nueve o diez veces menor que el nivel de expresión del marcador en una muestra de control (por ejemplo, muestra de sujetos sanos que no tienen el marcador asociado a un estado de enfermedad o un estado de toxicidad, por ejemplo, cáncer, diabetes, obesidad, enfermedad cardiovascular y cardiotoxicidad) y, preferiblemente, el nivel de expresión promedio del marcador en varias muestras de control.

Un “polinucleótido transcrito” o “transcripción de nucleótido” es un polinucleótido (por ejemplo, un ARNm, ARNh, un ADNc o un análogo de tal ARN o ADNc) que es complementario u homólogo con la totalidad o una parte de un ARNm maduro producido mediante la transcripción de un marcador de la invención y el procesamiento postranscripcional normal (por ejemplo, corte y empalme), si lo hubiera, del transcrito de ARN, y la transcripción inversa del transcrito de ARN.

“Complementario” se refiere al amplio concepto de complementariedad de secuencias entre regiones de dos hebras de ácido nucleico o entre dos regiones de la misma hebra de ácido nucleico. Se sabe que un residuo de adenina de una primera región de ácido nucleico es capaz de formar enlaces de hidrógeno específicos (“apareamiento de bases”) con un residuo de una segunda región de ácido nucleico que es antiparalela a la primera región si el residuo es timina o uracilo. De manera similar, se sabe que un residuo de citosina de una primera hebra de ácido nucleico es

capaz de aparearse con un resto de una segunda hebra de ácido nucleico que es antiparalela a la primera hebra si el residuo es guanina. Una primera región de un ácido nucleico es complementaria a una segunda región del mismo ácido nucleico o uno diferente si, cuando las dos regiones están dispuestas de manera antiparalela, al menos un residuo de nucleótido de la primera región es capaz de producir apareamiento de base con un residuo de la segunda región. Preferiblemente, la primera región comprende una primera parte y la segunda región comprende una segunda parte, mediante lo cual, cuando las partes primera y segunda están dispuestas de manera antiparalela, al menos aproximadamente el 50%, y preferiblemente al menos aproximadamente el 75%, al menos aproximadamente el 90%, o al menos aproximadamente el 95% de los residuos de nucleótido de la primera parte son capaces de producir apareamiento de bases con residuos de nucleótido en la segunda parte. Más preferiblemente, todos los residuos de nucleótido de la primera parte son capaces de producir apareamiento de bases con residuos de nucleótido en la segunda parte.

“Homólogo”, tal como se usa en el presente documento, se refiere a la similitud de secuencia de nucleótidos entre dos regiones de la misma hebra de ácido nucleico o entre regiones de dos hebras de ácido nucleico diferentes. Cuando una posición de residuo de nucleótido en ambas regiones está ocupada por el mismo residuo de nucleótido, entonces las regiones son homólogas en esa posición. Una primera región es homóloga a una segunda región si al menos una posición de residuo de nucleótido de cada región está ocupada por el mismo residuo. La homología entre dos regiones se expresa en términos de la proporción de posiciones de residuo de nucleótido de las dos regiones que están ocupadas por el mismo residuo de nucleótido. A modo de ejemplo, una región que tiene la secuencia de nucleótidos 5'-ATTGCC-3' y una región que tiene la secuencia de nucleótidos 5'-TATGGC-3' comparten una homología del 50%. Preferiblemente, la primera región comprende una primera parte y la segunda región comprende una segunda parte, mediante lo cual, al menos aproximadamente el 50%, y preferiblemente al menos aproximadamente el 75%, al menos aproximadamente el 90%, o al menos aproximadamente el 95% de las posiciones de residuo de nucleótido de cada una de las partes están ocupadas por el mismo residuo de nucleótido. Más preferiblemente, todas las posiciones de residuo de nucleótido de cada una de las partes están ocupadas por el mismo residuo de nucleótido.

Las “proteínas de la invención” abarcan proteínas marcadoras y sus fragmentos; proteínas marcadoras variantes y sus fragmentos; péptidos y polipéptidos que comprenden un segmento de al menos 15 aminoácidos de una proteína marcadora o proteína marcadora variante; y proteínas de fusión que comprenden una proteína marcadora o proteína marcadora variante, o un segmento de al menos 15 aminoácidos de una proteína marcadora o proteína marcadora variante.

La invención proporciona además anticuerpos, derivados de anticuerpo y fragmentos de anticuerpo que se unen específicamente con las proteínas marcadoras y los fragmentos de las proteínas marcadoras de la presente invención. A menos que se especifique lo contrario en el presente documento, los términos “anticuerpo” y “anticuerpos” abarcan ampliamente las formas de anticuerpos que se producen de manera natural (por ejemplo, IgG, IgA, IgM, IgE) y anticuerpos recombinantes tales como anticuerpos de cadena sencilla, anticuerpos quiméricos y humanizados y anticuerpos multiespecíficos, así como fragmentos y derivados de todos los anteriores, fragmentos y derivados que tienen al menos un sitio de unión a antígeno. Los derivados de anticuerpo pueden comprender un resto de proteína o químico conjugado con un anticuerpo.

En determinadas realizaciones, los marcadores de la invención incluyen uno o más genes (o proteínas) seleccionados del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1, CANX, GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. En algunas realizaciones, los marcadores son una combinación de al menos dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte, veinticinco, treinta o más de los genes (o proteínas) anteriores. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 de los genes (o proteínas) anteriores.

En una realización, los marcadores de la invención son genes o proteínas asociados al cáncer o implicados en el mismo. Tales genes o proteínas implicados en el cáncer incluyen, por ejemplo, HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y/o CANX. En algunas realizaciones, los marcadores de la invención son una combinación de al menos dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte o más de los genes (o proteínas) anteriores. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 de los genes (o proteínas) anteriores.

En una realización, los marcadores de la invención son genes o proteínas asociados a o implicados en la toxicidad inducida por fármacos. Tales genes o proteínas implicados en la toxicidad inducida por fármacos incluyen, por ejemplo, GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y/o TAZ. En algunas realizaciones, los marcadores de la invención son una combinación de al menos dos, tres, cuatro, cinco, seis, siete, ocho, nueve,

diez de los genes (o proteínas) anteriores. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 1 y 20, 1 y 30, 2 y 5, 2 y 10, 5 y 10, 1 y 20, 5 y 20, 10 y 20, 10 y 25, 10 y 30 de los genes (o proteínas) anteriores.

5 A. Marcadores asociados a cardiotoxicidad

La presente invención se basa, al menos en parte, en la identificación de nuevos biomarcadores que están asociados a cardiotoxicidad inducida por fármacos. La invención se basa además, al menos en parte, en el descubrimiento de que la coenzima Q10 es capaz de reducir o prevenir la cardiotoxicidad inducida por fármacos.

10 Por consiguiente, la invención proporciona métodos para identificar un agente que provoca o está en riesgo de provocar toxicidad. En una realización, el agente es un fármaco o candidato a fármaco. En una realización, la toxicidad es toxicidad inducida por fármacos, por ejemplo, cardiotoxicidad. En una realización, el agente es un fármaco o candidato a fármaco para tratar diabetes, obesidad o un trastorno cardiovascular. En estos métodos, se evalúa la cantidad de uno o más biomarcadores/proteínas en un par de muestras (una primera muestra no sujeta al tratamiento farmacológico y una segunda muestra sometida al tratamiento farmacológico). Una modulación del nivel de expresión de uno o más biomarcadores en la segunda muestra en comparación con la primera muestra es una indicación de que el fármaco provoca o está en riesgo de provocar toxicidad inducida por fármacos, por ejemplo, cardiotoxicidad. En una realización, el uno o más biomarcadores se selecciona del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. Los métodos de la presente invención pueden ponerse en práctica junto con cualquier otro método usado por el profesional experto para identificar un fármaco en riesgo de provocar cardiotoxicidad inducida por el fármaco.

25 Por consiguiente, en un aspecto, la invención proporciona un método para identificar un fármaco que provoca o está en riesgo de provocar toxicidad inducida por fármacos (por ejemplo, cardiotoxicidad), que comprende: comparar (i) el nivel de expresión del uno o más biomarcadores presentes en una primera muestra celular obtenida antes del tratamiento con el fármaco; con (ii) el nivel de expresión del uno o más biomarcadores presentes en una segunda muestra celular obtenida después del tratamiento con el fármaco; en el que el uno o más biomarcadores se seleccionan del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ; en el que una modulación del nivel de expresión del uno o más biomarcadores en la segunda muestra en comparación con la primera muestra es una indicación de que el fármaco provoca o está en riesgo de provocar toxicidad inducida por fármacos (por ejemplo, cardiotoxicidad).

35 En una realización, la toxicidad inducida por fármacos es cardiotoxicidad inducida por fármacos. En una realización, las células son células del sistema cardiovascular, por ejemplo, cardiomiocitos. En una realización, las células son cardiomiocitos de diabéticos. En una realización, el fármaco es un fármaco o fármaco candidato para tratar diabetes, obesidad o enfermedad cardiovascular.

40 En una realización, una modulación (por ejemplo, un aumento o una disminución) del nivel de expresión de uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o los diez biomarcadores seleccionados del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ en la segunda muestra en comparación con la primera muestra es una indicación de que el fármaco provoca o está en riesgo de provocar toxicidad inducida por fármacos.

45 La invención también proporciona métodos para identificar un agente que puede reducir o prevenir la toxicidad inducida por fármacos. En una realización, la toxicidad inducida por fármacos es cardiotoxicidad. En una realización, el fármaco es un fármaco o candidato a fármaco para tratar diabetes, obesidad o un trastorno cardiovascular. En estos métodos, se evalúa la cantidad de uno o más biomarcadores en tres muestras (una primera muestra no sometida al tratamiento farmacológico, una segunda muestra sometida al tratamiento farmacológico y una tercera muestra sometida tanto al tratamiento farmacológico como al agente). Aproximadamente el mismo nivel de expresión del uno o más biomarcadores en la tercera muestra en comparación con la primera muestra es una indicación de que el agente puede reducir o prevenir la toxicidad inducida por fármacos, por ejemplo, la cardiotoxicidad inducida por fármacos. En una realización, el uno o más biomarcadores se selecciona del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ.

55 Usando los métodos descritos en el presente documento, puede seleccionarse una variedad de moléculas, en particular moléculas que son lo suficientemente pequeñas como para poder atravesar la membrana celular, con el fin de identificar moléculas que modulan, por ejemplo, aumentan o disminuyen la expresión y/o actividad de un marcador de la invención. Los compuestos así identificados pueden proporcionarse a un sujeto para reducir, aliviar o prevenir la toxicidad inducida por fármacos en el sujeto.

60 Por consiguiente, en otro aspecto, la invención proporciona un método para identificar un agente que puede reducir o prevenir la toxicidad inducida por fármacos, que comprende: (i) determinar el nivel de expresión del uno o más biomarcadores presentes en una primera muestra celular obtenida antes del tratamiento con un fármaco inductor de toxicidad; (ii) determinar el nivel de expresión del uno o más biomarcadores presentes en una segunda muestra celular obtenida después del tratamiento con el fármaco inductor de toxicidad; (iii) determinar el nivel de expresión

del uno o más biomarcadores presentes en una tercera muestra celular obtenida después del tratamiento con el fármaco inductor de toxicidad y el agente; y (iv) comparar el nivel de expresión del uno o más biomarcadores presentes en la tercera muestra con la primera muestra; en el que el uno o más biomarcadores se seleccionan del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ; y en el que aproximadamente el mismo nivel de expresión del uno o más biomarcadores en la tercera muestra en comparación con la primera muestra es una indicación de que el agente puede reducir o prevenir la toxicidad inducida por fármacos.

En una realización, la toxicidad inducida por fármacos es cardiotoxicidad inducida por fármacos. En una realización, las células son células del sistema cardiovascular, por ejemplo, cardiomiocitos. En una realización, las células son cardiomiocitos de diabéticos. En una realización, el fármaco es un fármaco o fármaco candidato para tratar diabetes, obesidad o enfermedad cardiovascular.

En una realización, aproximadamente el mismo nivel de expresión de uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o los diez biomarcadores seleccionados del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ en la tercera muestra en comparación con la primera muestra es una indicación de que el agente puede reducir o prevenir la toxicidad inducida por fármacos.

La invención proporciona además métodos para aliviar, reducir o prevenir la cardiotoxicidad inducida por fármacos en un sujeto que lo necesita, que comprende administrar a un sujeto (por ejemplo, un mamífero, un humano o un animal no humano) un agente identificado mediante los métodos de examen proporcionados en el presente documento, reduciendo o previniendo así la cardiotoxicidad inducida por fármacos en el sujeto. En una realización, el agente se administra a un sujeto que ya se ha tratado con un fármaco inductor de cardiotoxicidad. En una realización, el agente se administra a un sujeto al mismo tiempo que el tratamiento del sujeto con un fármaco inductor de cardiotoxicidad. En una realización, el agente se administra a un sujeto antes del tratamiento del sujeto con un fármaco inductor de cardiotoxicidad.

La invención proporciona además métodos para aliviar, reducir o prevenir la cardiotoxicidad inducida por fármacos en un sujeto que lo necesita, que comprende administrar coenzima Q10 al sujeto (por ejemplo, un mamífero, un humano o un animal no humano), reduciendo o previniendo de ese modo la cardiotoxicidad inducida por fármacos en el sujeto. En una realización, la coenzima Q10 se administra a un sujeto que ya se ha tratado con un fármaco inductor de cardiotoxicidad. En una realización, la coenzima Q10 se administra a un sujeto al mismo tiempo que el tratamiento del sujeto con un fármaco inductor de cardiotoxicidad. En una realización, la coenzima Q10 se administra a un sujeto antes del tratamiento del sujeto con un fármaco inductor de cardiotoxicidad. En una realización, la cardiotoxicidad inducida por fármacos está asociada con la modulación de la expresión de uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o los diez biomarcadores seleccionados del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. Todos los valores presentados en la lista anterior también pueden ser el límite superior o inferior de intervalos, que pretenden formar parte de esta invención, por ejemplo, entre 1 y 5, 1 y 10, 2 y 5, 2 y 10, o 5 y 10 de los genes (o proteínas) anteriores.

La invención proporciona además biomarcadores (por ejemplo, genes y/o proteínas) que son útiles como marcadores predictivos de cardiotoxicidad, por ejemplo, cardiotoxicidad inducida por fármacos. Estos biomarcadores incluyen GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. Sin embargo, el experto en la técnica será capaz de identificar biomarcadores adicionales predictivos de cardiotoxicidad inducida por fármacos empleando los métodos descritos en el presente documento, por ejemplo, llevando a cabo los métodos descritos en el ejemplo 3 pero usando un fármaco diferente que se sabe que induce cardiotoxicidad. Los ejemplos de biomarcadores de cardiotoxicidad inducida por fármacos de la invención se describen adicionalmente a continuación.

GRP78 y GRP75 también se denominan proteínas de respuesta a la glucosa. Estas proteínas están asociadas con el estrés del retículo endo/sarcoplasmático (estrés de RE) de los cardiomiocitos. SERCA, o ATPasa de calcio del retículo sarcoendoplasmático, regula la homeostasis de Ca^{2+} en células cardíacas. Cualquier perturbación de estas ATPasas puede conducir a disfunción cardíaca e insuficiencia cardíaca. Basándose en los datos proporcionados en el presente documento, GRP75 y GRP78 y las aristas a su alrededor son nuevos factores pronóstico de cardiotoxicidad inducida por fármacos.

TIMP1, también denominado inhibidor 1 de la metaloproteasa TIMP, participa en la remodelación de la matriz extracelular en asociación con las MMP. La expresión de TIMP1 se correlaciona con fibrosis del corazón, y la hipoxia de las células endoteliales vasculares también induce la expresión de TIMP1. Basándose en los datos proporcionados en el presente documento, TIMP1 es un nuevo factor pronóstico de la toxicidad cardíaca inducida por fármacos.

PTX3, también denominado pentraxina 3, pertenece a la familia de las proteínas C reactivas (PCR) y es un buen marcador de un estado inflamatorio del corazón. Sin embargo, la PTX3 plasmática también podría ser representativa de la respuesta inflamatoria sistémica debida a septicemia u otros estados médicos. Basándose en los datos proporcionados en el presente documento, PTX3 puede ser un nuevo marcador de la función cardíaca o

cardiotoxicidad. Además, las aristas asociadas con PTX3 en la red podrían formar un nuevo panel de biomarcadores.

5 HSP76, también denominado HSPA6, sólo se sabe que se expresa en células endoteliales y linfocitos B. No se conoce el papel de esta proteína en la función cardíaca. Basándose en los datos proporcionados en el presente documento, HSP76 puede ser un nuevo factor pronóstico de cardiotoxicidad inducida por fármacos.

10 PDIA4, PDIA1, también denominados proteínas de la familia A de proteína disulfuro isomerasas, están asociadas con la respuesta al estrés de RE, como las GRP. No se conoce el papel de estas proteínas en la función cardíaca. Basándose en los datos proporcionados en el presente documento, estas proteínas pueden ser nuevos factores pronóstico de cardiotoxicidad inducida por fármacos.

15 CA2D1 también denominado subunidad alfa 2/delta dependiente de voltaje, de canal de calcio. La subunidad alfa-2/delta del canal de calcio dependiente de voltaje regula la densidad de corriente de calcio y la cinética de activación/inactivación del canal de calcio. CA2D1 desempeña un papel importante en el acoplamiento de excitación-contracción en el corazón. No se conoce el papel de esta proteína en la función cardíaca. Basándose en los datos proporcionados en el presente documento, CA2D1 es un nuevo factor pronóstico de cardiotoxicidad inducida por fármacos.

20 GPAT1 es una de las cuatro isoformas conocidas de glicerol-3-fosfato aciltransferasa, y se encuentra en la membrana externa mitocondrial, permitiendo la regulación recíproca con carnitina palmitoiltransferasa-1. GPAT1 está regulado por incremento a nivel transcripcional por insulina y SREBP-1c y regulado por disminución de forma aguda por la proteína cinasa activada por AMP, lo que concuerda con un papel en la síntesis de triacilglicerol. Basándose en los datos proporcionados en el presente documento, GPAT1 es un nuevo factor pronóstico de cardiotoxicidad inducida por fármacos.

30 TAZ, también denominado tafazzina, se expresa altamente en el músculo cardíaco y esquelético. TAZ está implicada en el metabolismo de la cardiolipina y funciona como fosfolípido-lisofosfolípido transacilasa. La tafazzina es responsable de la remodelación de un fosfolípido cardiolipina (CL), el lípido característico de la membrana interna mitocondrial. Basándose en los datos proporcionados en el presente documento, TAZ es un nuevo factor pronóstico de cardiotoxicidad inducida por fármacos.

B. Marcadores asociados a cáncer

35 La presente invención se basa, al menos en parte, en la identificación de nuevos biomarcadores que están asociados con el cáncer. Tales marcadores asociados en el cáncer incluyen, por ejemplo, HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y/o CANX. En algunas realizaciones, los marcadores de la invención son una combinación de al menos dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte o más de los marcadores anteriores.

45 Por consiguiente, la invención proporciona métodos para identificar un agente que provoca o está en riesgo de provocar cáncer. En una realización, el agente es un fármaco o candidato a fármaco. En estos métodos, se evalúa la cantidad de uno o más biomarcadores/proteínas en un par de muestras (una primera muestra no sujeta al tratamiento farmacológico y una segunda muestra sometida al tratamiento farmacológico). Una modulación del nivel de expresión de uno o más biomarcadores en la segunda muestra en comparación con la primera muestra es una indicación de que el fármaco provoca o está en riesgo de provocar cáncer. En una realización, el uno o más biomarcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX. Los métodos de la presente invención pueden ponerse en práctica junto con cualquier otro método usado por el profesional experto para identificar un fármaco en riesgo de provocar el cáncer.

55 En un aspecto, la invención proporciona métodos para evaluar la eficacia de una terapia para tratar un cáncer en un sujeto, comprendiendo el método: comparar el nivel de expresión del uno o más marcadores presentes en una primera muestra obtenida del sujeto antes de administrar al menos una parte del régimen de tratamiento al sujeto, en el que el uno o más marcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX; y el nivel de expresión del uno o más marcadores presentes en una segunda muestra obtenida del sujeto después de la administración de al menos una parte del régimen de tratamiento, en el que una modulación del nivel de expresión del uno o más marcadores en la segunda muestra en comparación con la primera muestra es una indicación de que la terapia es eficaz para tratar el cáncer en el sujeto.

65 En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.

- 5 En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.
- 5 En una realización, el sujeto es un humano.
- 10 En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.
- 15 En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.
- 20 En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.
- 25 En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.
- En una realización, se determina el nivel de expresión de una pluralidad de marcadores.
- 30 En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.
- 35 La invención proporciona además métodos para evaluar si un sujeto está afectado por un cáncer, comprendiendo el método: determinar el nivel de expresión del uno o más marcadores presentes en una muestra biológica obtenida del sujeto, en el que el uno o más marcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX; y comparar el nivel de expresión del uno o más marcadores presentes en la muestra biológica obtenida del sujeto con el nivel de expresión del uno o más marcadores presentes en una muestra de control, en el que una modulación del nivel de expresión del uno o más marcadores en la muestra biológica obtenida del sujeto en relación con el nivel de expresión del uno o más marcadores en la muestra de control es una indicación de que el sujeto está afectado por cáncer, evaluando así si el sujeto está afectado por el cáncer.
- 45 En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.
- 50 En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.
- 55 En una realización, el sujeto es un humano.
- En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.
- 60 En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.
- 65 En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa

- 5 (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.
- 10 En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.
- 15 En una realización, se determina el nivel de expresión de una pluralidad de marcadores.
- 20 En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.
- 25 La invención proporciona además métodos para pronosticar si un sujeto está predispuesto a desarrollar un cáncer, comprendiendo el método: determinar el nivel de expresión de uno o más marcadores presentes en una muestra biológica obtenida del sujeto, en el que el uno o más marcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX; y comparar el nivel de expresión del uno o más marcadores presentes en la muestra biológica obtenida del sujeto con el nivel de expresión del uno o más marcadores presentes en una muestra de control, en el que una modulación del nivel de expresión del uno o más marcadores en la muestra biológica obtenida del sujeto en relación con el nivel de expresión del uno o más marcadores en la muestra de control es una indicación de que el sujeto está predispuesto a desarrollar cáncer, pronosticándose de ese modo si el sujeto está predispuesto a desarrollar el cáncer.
- 30 En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.
- 35 En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.
- 40 En una realización, el sujeto es un humano.
- 45 En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.
- 50 En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.
- 55 En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.
- 60 En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.
- 65 En una realización, se determina el nivel de expresión de una pluralidad de marcadores.
- En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.

La invención proporciona además métodos de pronóstico de la recidiva de un cáncer en un sujeto, comprendiendo el método: determinar el nivel de expresión de uno o más marcadores presentes en una muestra biológica obtenida del sujeto, en el que el uno o más marcadores se seleccionan de grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX; y comparar el nivel de expresión del uno o más marcadores presentes en la muestra biológica obtenida del sujeto con el nivel de expresión del uno o más marcadores presentes en una muestra de control, en el que una modulación del nivel de expresión del uno o más marcadores en la muestra biológica obtenida del sujeto en relación con el nivel de expresión del uno o más marcadores en la muestra de control es una indicación de la recidiva del cáncer, pronosticándose de ese modo la recidiva del cáncer en el sujeto.

En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.

En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.

En una realización, el sujeto es un humano.

En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.

En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.

En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.

En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.

En una realización, se determina el nivel de expresión de una pluralidad de marcadores.

En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.

Además, la invención proporciona métodos de pronóstico de la supervivencia de un sujeto con cáncer, comprendiendo el método: determinar el nivel de expresión de uno o más marcadores presentes en una muestra biológica obtenida del sujeto, en el que el uno o más marcadores se seleccionan de grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX; y comparar el nivel de expresión del uno o más marcadores presentes en la muestra biológica obtenida del sujeto con el nivel de expresión del uno o más marcadores presentes en una muestra de control, en el que una modulación del nivel de expresión del uno o más marcadores en la muestra biológica obtenida del sujeto en relación con el nivel de expresión del uno o más marcadores en la muestra de control es una indicación de la supervivencia del sujeto, pronosticándose de ese modo la supervivencia del sujeto con el cáncer.

En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.

En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una

realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.

En una realización, el sujeto es un humano.

5 En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.

10 En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.

15 En una realización, el nivel de expresión de uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.

20 En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.

25 En una realización, se determina el nivel de expresión de una pluralidad de marcadores.

30 En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.

35 La invención proporciona además métodos de monitorización de la progresión de un cáncer en un sujeto, comprendiendo el método: comparar el nivel de expresión de uno o más marcadores presentes en una primera muestra obtenida del sujeto antes de administrar al menos una parte de un régimen de tratamiento al sujeto y el nivel de expresión del uno o más marcadores presentes en una segunda muestra obtenida del sujeto después de la administración de al menos una parte del régimen de tratamiento, en el que el uno o más marcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX, monitorizándose de ese modo la progresión del cáncer en el sujeto.

40 En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.

45 En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.

50 En una realización, el sujeto es un humano.

55 En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.

60 En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.

65 En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de

fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.

En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.

En una realización, se determina el nivel de expresión de una pluralidad de marcadores.

En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.

La invención proporciona además métodos de identificación de un compuesto para tratar un cáncer en un sujeto, comprendiendo el método: obtener una muestra biológica del sujeto; poner en contacto la muestra biológica con un compuesto de prueba; determinar el nivel de expresión de uno o más marcadores presentes en la muestra biológica obtenida del sujeto, en el que el uno o más marcadores se seleccionan del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX con un cambio en veces positivo y/o con un cambio en veces negativo; comparar el nivel de expresión del uno o más marcadores en la muestra biológica con un control apropiado; y seleccionar un compuesto de prueba que disminuye el nivel de expresión del uno o más marcadores con un cambio en veces negativo presente en la muestra biológica y/o aumenta el nivel de expresión del uno o más marcadores con un cambio en veces positivo presente en la muestra biológica, identificándose de ese modo un compuesto para tratar el cáncer en un sujeto.

En una realización, la muestra comprende un líquido obtenido del sujeto. En una realización, el líquido se selecciona del grupo que consiste en líquidos sanguíneos, vómito, saliva, linfa, líquido quístico, orina, líquidos recogidos mediante lavado bronquial, líquidos recogidos mediante enjuague peritoneal y líquidos ginecológicos. En una realización, la muestra es una muestra de sangre o un componente de la misma.

En otra realización, la muestra comprende un tejido o componente del mismo obtenido del sujeto. En una realización, el tejido se selecciona del grupo que consiste en hueso, tejido conjuntivo, cartílago, pulmón, hígado, riñón, tejido muscular, corazón, páncreas y piel.

En una realización, el sujeto es un humano.

En una realización, el nivel de expresión del uno o más marcadores en la muestra biológica se determina sometiendo a ensayo un polinucleótido transcrito o una parte del mismo en la muestra. En una realización, en la que el ensayo del polinucleótido transcrito comprende amplificar el polinucleótido transcrito.

En una realización, el nivel de expresión del marcador en la muestra del sujeto se determina sometiendo a ensayo una proteína o una parte de la misma en la muestra. En una realización, la proteína se somete a ensayo usando un reactivo que se une específicamente con la proteína.

En una realización, el nivel de expresión del uno o más marcadores en la muestra se determina usando una técnica seleccionada del grupo que consiste en la reacción de amplificación de la reacción en cadena de la polimerasa (PCR), análisis por PCR con transcriptasa inversa, análisis de polimorfismos de conformación de cadena sencilla (SSCP), detección de escisión de apareamiento erróneo, análisis de heterodúplex, análisis de transferencia de tipo Southern, análisis de transferencia de tipo Northern, análisis de inmunotransferencia de tipo Western, hibridación *in situ*, análisis de alineamiento, secuenciación de ácido desoxirribonucleico, análisis de polimorfismos de longitud de fragmento de restricción, y combinaciones o subcombinaciones de los mismos, de dicha muestra.

En una realización, el nivel de expresión del marcador en la muestra se determina usando una técnica seleccionada del grupo que consiste en inmunohistoquímica, inmunocitoquímica, citometría de flujo, ELISA y espectrometría de masas.

En una realización, se determina el nivel de expresión de una pluralidad de marcadores.

En una realización, el sujeto está tratándose con una terapia seleccionada del grupo que consiste en un compuesto influyente ambiental, cirugía, radiación, terapia hormonal, terapia con anticuerpos, terapia con factores de crecimiento, citocinas, quimioterapia, terapia con células madre alogénicas. En una realización, el compuesto influyente ambiental es una molécula de coenzima Q10.

La invención proporciona además un kit para evaluar la eficacia de una terapia para tratar un cáncer, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5,

DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para evaluar la eficacia de la terapia para tratar el cáncer.

5 La invención proporciona además un kit para evaluar si un sujeto padece un cáncer, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para evaluar si el sujeto está afectado por el cáncer.

10 La invención proporciona además un kit para pronosticar si un sujeto está predispuesto a desarrollar un cáncer, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para pronosticar si el sujeto está predispuesto a desarrollar el cáncer.

15 La invención proporciona además un kit para pronosticar la recidiva de un cáncer en un sujeto, comprendiendo el kit reactivos para evaluar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para pronosticar la recidiva del cáncer.

20 La invención proporciona además un kit para pronosticar la recidiva de un cáncer, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para pronosticar la recidiva del cáncer.

25 La invención proporciona además un kit para pronosticar la supervivencia de un sujeto con cáncer, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para pronosticar la supervivencia del sujeto con el cáncer.

30 La invención proporciona además un kit para monitorizar la progresión de un cáncer en un sujeto, comprendiendo el kit reactivos para determinar el nivel de expresión de al menos un marcador seleccionado del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX e instrucciones para el uso del kit para pronosticar la progresión del cáncer en un sujeto.

35 Los kits de la invención pueden comprender además medios para obtener una muestra biológica de un sujeto, una muestra de control y/o un compuesto influyente ambiental.

40 Los medios para determinar el nivel de expresión de al menos un marcador pueden comprender medios para someter a ensayo un polinucleótido transcrito o una parte del mismo en la muestra y/o medios para someter a ensayo una proteína o una parte del mismo en la muestra.

45 En una realización, los kits comprenden reactivos para determinar el nivel de expresión de una pluralidad de marcadores.

50 Diversos aspectos de la invención se describen con más detalle en las siguientes subsecciones.

C. Moléculas de ácido nucleico aisladas

55 Un aspecto de la invención se refiere a moléculas de ácido nucleico aisladas, incluyendo ácidos nucleicos que codifican para una proteína marcadora o una parte de la misma. Los ácidos nucleicos aislados de la invención también incluyen moléculas de ácido nucleico suficientes para su uso como sondas de hibridación para identificar moléculas de ácido nucleico marcador, y fragmentos de moléculas de ácido nucleico marcador, por ejemplo, aquellas adecuadas para su uso como cebadores de PCR para la amplificación o mutación de moléculas de ácido nucleico marcador. Tal como se usa en el presente documento, el término "molécula de ácido nucleico" pretende incluir moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm) y análogos del ADN o ARN generado usando análogos de nucleótidos. La molécula de ácido nucleico puede ser monocatenaria o bicatenaria, pero preferiblemente es ADN bicatenario.

60 Una molécula de ácido nucleico "aislada" es una que está separada de otras moléculas de ácido nucleico que están presentes en la fuente natural de la molécula de ácido nucleico. En una realización, una molécula de ácido nucleico "aislada" está libre de secuencias (preferiblemente secuencias que codifican para proteínas) que flanquean de

manera natural al ácido nucleico (es decir, secuencias ubicadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del que deriva el ácido nucleico. Por ejemplo, en diversas realizaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0,5 kB o 0,1 kB de secuencias de nucleótidos que flanquean de manera natural a la molécula de ácido nucleico en el ADN genómico de la célula de la que deriva el ácido nucleico. En otra realización, una molécula de ácido nucleico "aislada", tal como una molécula de ADNc, puede estar sustancialmente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente. Una molécula de ácido nucleico que está sustancialmente libre de material celular incluye preparaciones que tienen menos de aproximadamente el 30%, el 20%, el 10% o el 5% de ácido nucleico heterólogo (también denominado en el presente documento "ácido nucleico contaminante").

Una molécula de ácido nucleico de la presente invención puede aislarse usando técnicas de biología molecular convencionales y la información de secuencia en los registros de base de datos descritos en el presente documento. Usando la totalidad o una parte de tales secuencias de ácido nucleico, pueden aislarse moléculas de ácido nucleico de la invención usando técnicas de hibridación y clonación convencionales (por ejemplo, tal como se describe en Sambrook *et al.*, Ed., *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Una molécula de ácido nucleico de la invención puede amplificarse usando ADNc, ARNm o ADN genómico como molde y cebadores oligonucleotídicos apropiados según técnicas de amplificación de PCR convencionales. El ácido nucleico así amplificado puede clonarse en un vector apropiado y caracterizarse mediante análisis de secuencia de ADN. Además, pueden prepararse nucleótidos correspondientes a la totalidad o una parte de una molécula de ácido nucleico de la invención mediante técnicas de síntesis convencionales, por ejemplo, usando un sintetizador de ADN automatizado.

En otra realización preferida, una molécula de ácido nucleico aislada de la invención comprende una molécula de ácido nucleico que tiene una secuencia de nucleótidos complementaria a la secuencia de nucleótidos de un ácido nucleico marcador o a la secuencia de nucleótidos de un ácido nucleico que codifica para una proteína marcadora. Una molécula de ácido nucleico que es complementaria a una secuencia de nucleótidos dada es una que es suficientemente complementaria a la secuencia de nucleótidos dada que puede hibridarse con la secuencia de nucleótidos dada formando así un dúplex estable.

Además, una molécula de ácido nucleico de la invención puede comprender sólo una parte de una secuencia de ácido nucleico, en la que la secuencia de ácido nucleico de longitud completa comprende un ácido nucleico marcador o que codifica para una proteína marcadora. Tales ácidos nucleicos pueden usarse, por ejemplo, como sonda o cebador. La sonda/cebador se usa normalmente como uno o más oligonucleótidos sustancialmente purificados. El oligonucleótido comprende normalmente una región de secuencia de nucleótidos que se hibrida en condiciones rigurosas con al menos aproximadamente 7, de manera aproximada preferiblemente 15, de manera aproximada más preferiblemente 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 ó 400 o más nucleótidos consecutivos de un ácido nucleico de la invención.

Las sondas basadas en la secuencia de una molécula de ácido nucleico de la invención pueden usarse para detectar transcritos o secuencias genómicas correspondientes a uno o más marcadores de la invención. La sonda comprende un grupo de etiqueta unido a la misma, por ejemplo, un radioisótopo, un compuesto fluorescente, una enzima o un cofactor enzimático. Tales sondas pueden usarse como parte de un kit de prueba de diagnóstico para identificar células o tejidos que expresan erróneamente la proteína, por ejemplo midiendo los niveles de una molécula de ácido nucleico que codifica para la proteína en una muestra de células de un sujeto, por ejemplo, detectando niveles de ARNm o determinando si un gen que codifica para la proteína se ha mutado o deletado.

La invención abarca además moléculas de ácido nucleico que difieren, debido a la degeneración del código genético, de la secuencia de nucleótidos de los ácidos nucleicos que codifican para una proteína marcadora y, por tanto, codifican para la misma proteína.

Los expertos en la técnica apreciarán que pueden existir polimorfismos de secuencia de ADN que conducen a cambios en la secuencia de aminoácidos dentro de una población (por ejemplo, la población humana). Tales polimorfismos genéticos pueden existir entre individuos dentro de una población debido a la variación alélica natural. Un alelo es uno de un grupo de genes que aparecen alternativamente en un locus genético dado. Además, se apreciará que también pueden existir polimorfismos de ADN que afectan a los niveles de expresión de ARN, lo que pueden afectar al nivel de expresión global de ese gen (por ejemplo, afectando a la regulación o la degradación).

Tal como se usa en el presente documento, la expresión "variante alélica" se refiere a una secuencia de nucleótidos que aparece en un locus dado o a un polipéptido codificado por la secuencia de nucleótidos.

Tal como se usan en el presente documento, los términos "gen" y "gen recombinante" se refieren a moléculas de ácido nucleico que comprenden un marco de lectura abierto que codifica para un polipéptido correspondiente a un marcador de la invención. Tales variaciones alélicas naturales pueden dar como resultado una variación del 1-5% en

la secuencia de nucleótidos de un gen dado. Pueden identificarse alelos alternativos secuenciando el gen de interés en varios individuos diferentes. Esto puede llevarse a cabo fácilmente mediante el uso de sondas de hibridación para identificar el mismo locus genético en una variedad de individuos. Se pretende que todas y cada una de estas variaciones de nucleótidos y polimorfismos de aminoácidos resultantes o variaciones que son el resultado de la variación alélica natural y que no alteran la actividad funcional estén dentro del alcance de la invención.

En otra realización, una molécula de ácido nucleico aislada de la invención tiene al menos 7, 15, 20, 25, 30, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500 o más nucleótidos de longitud y se hibrida en condiciones rigurosas con un ácido nucleico marcador o con un ácido nucleico que codifica para una proteína marcadora. Tal como se usa en el presente documento, el término "se hibrida en condiciones rigurosas" pretende describir condiciones para la hibridación y el lavado en las que las secuencias de nucleótidos idénticas en al menos el 60% (el 65%, el 70%, preferiblemente el 75%) entre sí permanecen normalmente hibridadas entre sí. Tales condiciones rigurosas las conocen los expertos en la técnica y pueden hallarse en las secciones 6.3.1-6.3.6 de *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989)). Un ejemplo preferido y no limitativo de condiciones de hibridación rigurosas es la hibridación en cloruro de sodio 6X/citrato de sodio (SSC) a aproximadamente 45°C, seguido de uno o más lavados en SSC 0,2X, SDS al 0,1% a 50-65°C.

Además de las variantes alélicas que se producen de manera natural de una molécula de ácido nucleico de la invención que pueden existir en la población, el experto en la técnica apreciará que pueden introducirse cambios de secuencia mediante mutación, lo que conduce a cambios en la secuencia de aminoácidos de la proteína codificada, sin alterar la actividad biológica de la proteína codificada por la misma. Por ejemplo, pueden realizarse sustituciones de nucleótidos que conducen a sustituciones de aminoácidos en residuos de aminoácido "no esenciales". Un residuo de aminoácido "no esencial" es un residuo que puede alterarse con respecto a la secuencia de tipo natural sin alterar la actividad biológica, mientras que se requiere un residuo de aminoácido "esencial" para la actividad biológica. Por ejemplo, los residuos de aminoácido que no están conservados o sólo semiconservados entre homólogos de diversas especies pueden no ser esenciales para la actividad y, por tanto, podrían ser dianas para la alteración. Alternativamente, los residuos de aminoácido que se conservan entre los homólogos de diversas especies (por ejemplo, murino y humano) pueden ser esenciales para la actividad y, por tanto, no serían dianas probables para la alteración.

Por consiguiente, otro aspecto de la invención se refiere a moléculas de ácido nucleico que codifican para una proteína marcadora variante que contiene cambios en los residuos de aminoácido que no son esenciales para la actividad. Tales proteínas marcadoras variantes difieren en la secuencia de aminoácidos de las proteínas marcadoras que se producen de manera natural, pero conservan la actividad biológica. En una realización, tal proteína marcadora variante tiene una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 40%, el 50%, el 60%, el 70%, el 80%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99% a la secuencia de aminoácidos de una proteína marcadora.

Puede crearse una molécula de ácido nucleico aislada que codifique una proteína marcadora variante mediante la introducción de una o más sustituciones, adiciones o deleciones de nucleótidos en la secuencia de nucleótidos de los ácidos nucleicos marcadores, de tal manera que se introduzcan una o más sustituciones, adiciones o deleciones de residuos de aminoácido en la proteína codificada. Pueden introducirse mutaciones mediante técnicas convencionales, tales como mutagénesis dirigida al sitio y mutagénesis mediada por PCR. Preferiblemente, se realizan sustituciones conservativas de aminoácidos en uno o más residuos de aminoácido no esenciales predichos. Una "sustitución conservativa de aminoácidos" es aquella en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. Las familias de residuos de aminoácido que tienen cadenas laterales similares se han definido en la técnica. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales apolares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales beta-ramificadas (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Alternativamente, pueden introducirse mutaciones aleatoriamente a lo largo de la totalidad o parte de la secuencia codificante, tal como mediante mutagénesis de saturación, y los mutantes resultantes pueden examinarse en busca de actividad biológica para identificar mutantes que retienen la actividad. Después de la mutagénesis, la proteína codificada puede expresarse de manera recombinante y puede determinarse la actividad de la proteína.

La presente invención abarca moléculas de ácido nucleico antisentido, es decir, moléculas que son complementarias a un ácido nucleico sentido de la invención, por ejemplo, complementarias a la hebra codificante de una molécula de ADNc marcador bicatenario o complementarias a una secuencia de ARNm marcador. Por consiguiente, un ácido nucleico antisentido de la invención puede unirse con enlaces de hidrógeno a (es decir, aparearse con) un ácido nucleico sentido de la invención. El ácido nucleico antisentido puede ser complementario a una hebra codificante completa, o sólo a una parte de la misma, por ejemplo, la totalidad o parte de la región codificante de proteína (o marco de lectura abierto). Una molécula de ácido nucleico antisentido también puede ser antisentido para la totalidad o parte de una región no codificante de la hebra codificante de una secuencia de nucleótidos que codifica

para una proteína marcadora. Las regiones no codificantes ("regiones no traducidas en 5' y 3'") son las secuencias en 5' y 3' que flanquean a la región codificante y no se traducen en aminoácidos.

5 Un oligonucleótido antisentido puede tener, por ejemplo, aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50 o más nucleótidos de longitud. Puede construirse un ácido nucleico antisentido de la invención usando síntesis química y reacciones de ligamiento enzimático usando procedimientos conocidos en la técnica. Por ejemplo, un ácido nucleico antisentido (por ejemplo, un oligonucleótido antisentido) puede sintetizarse químicamente usando nucleótidos que se producen de manera natural o nucleótidos modificados de manera variada diseñados para aumentar la estabilidad biológica de las moléculas o para aumentar la estabilidad física del dúplex formado entre los ácidos nucleicos antisentido y sentido, por ejemplo, pueden usarse derivados de fosforotioato y nucleótidos sustituidos con acridina. Los ejemplos de nucleótidos modificados que pueden usarse para generar el ácido nucleico antisentido incluyen 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroxi)metiluracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, dihidouracilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiamino-metil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxycarboximetiluracilo, 5-metoxiuracilo, 2-metil-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil)uracilo, (acp3)w y 2,6-diaminopurina. Alternativamente, el ácido nucleico antisentido puede producirse biológicamente usando un vector de expresión en el que se ha subclonado un ácido nucleico en una orientación antisentido (es decir, el ARN transcrito a partir del ácido nucleico insertado tendrá una orientación antisentido con respecto a un ácido nucleico diana de interés, descrito más adelante en la siguiente subsección).

25 Las moléculas de ácido nucleico antisentido de la invención se administran normalmente a un sujeto o se generan *in situ* de modo que se hibridan con o se unen al ARNm celular y/o al ADN genómico que codifica para una proteína marcadora para inhibir de ese modo la expresión del marcador, por ejemplo, inhibiendo la transcripción y/o traducción. La hibridación puede ser por complementariedad de nucleótidos convencional para formar un dúplex estable o, por ejemplo, en el caso de una molécula de ácido nucleico antisentido que se une a un dúplex de ADN, a través de interacciones específicas en el surco mayor de la doble hélice. Los ejemplos de una ruta de administración de moléculas de ácido nucleico antisentido de la invención incluyen inyección directa en un sitio de tejido o infusión del ácido nucleico antisentido en líquido corporal asociado a estado de enfermedad o estado de toxicidad. Alternativamente, las moléculas de ácido nucleico antisentido pueden modificarse para seleccionarse como diana células seleccionadas y luego administrarse de manera sistémica. Por ejemplo, para la administración sistémica, las moléculas antisentido pueden modificarse de modo que se unan específicamente a receptores o antígenos expresados en una superficie celular seleccionada, por ejemplo, ligando las moléculas de ácido nucleico antisentido a péptidos o anticuerpos que se unen a receptores o antígenos de la superficie celular. Las moléculas de ácido nucleico antisentido también pueden administrarse a las células usando los vectores descritos en el presente documento. Para lograr concentraciones intracelulares suficientes de las moléculas antisentido, se prefieren los constructos de vector en los que la molécula de ácido nucleico antisentido se coloca bajo el control de un promotor fuerte pol II o pol III.

Una molécula de ácido nucleico antisentido de la invención puede ser una molécula de ácido nucleico α -anomérico. Una molécula de ácido nucleico α -anomérico forma híbridos bicatenarios específicos con ARN complementario en el que, al contrario de las unidades α habituales, las hebras discurren paralelas entre sí (Gaultier *et al.*, 1987, Nucleic Acids Res. 15:6625-6641). La molécula de ácido nucleico antisentido también puede comprender un 2'-O-metil-ribonucleótido (Inoue *et al.*, 1987, Nucleic Acids Res. 15:6131-6148) o un análogo de ARN-ADN quimérico (Inoue *et al.*, 1987, FEBS Lett. 215:327-330).

50 La invención también abarca ribozimas. Las ribozimas son moléculas de ARN catalíticas con actividad ribonucleasa que son capaces de escindir un ácido nucleico monocatenario, tal como un ARNm, con el que tienen una región complementaria. Por tanto, las ribozimas (por ejemplo, ribozimas de cabeza de martillo tal como se describe en Haselhoff y Gerlach, 1988, Nature 334:585-591) pueden usarse para escindir catalíticamente transcritos de ARNm para inhibir de ese modo la traducción de la proteína codificada por el ARNm. Una ribozima que tiene especificidad por una molécula de ácido nucleico que codifica para una proteína marcadora puede diseñarse basándose en la secuencia de nucleótidos de un ADNc correspondiente al marcador. Por ejemplo, puede construirse un derivado de un ARN de L-19 IVS de *Tetrahymena* en el que la secuencia de nucleótidos del sitio activo es complementaria a la secuencia de nucleótidos que va a escindirse (véanse Cech *et al.* la patente estadounidense n.º 4.987.071; y Cech *et al.* la patente estadounidense n.º 5.116.742). Alternativamente, un ARNm que codifica para un polipéptido de la invención puede usarse para seleccionar un ARN catalítico que tiene una actividad ribonucleasa específica de una reserva de moléculas de ARN (véase, por ejemplo, Bartel y Szostak, 1993, Science 261:1411-1418).

65 La invención también abarca moléculas de ácido nucleico que forman estructuras helicoidales triples. Por ejemplo, la expresión de un marcador de la invención puede inhibirse seleccionando como diana secuencias de nucleótidos complementarias a la región reguladora del gen que codifica para la proteína o el ácido nucleico marcador (por

ejemplo, el promotor y/o potenciador) para formar estructuras helicoidales triples que impiden la transcripción del gen en las células diana. Véanse en general Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; y Maher (1992) *Bioassays* 14(12):807-15.

5 En diversas realizaciones, las moléculas de ácido nucleico de la invención pueden modificarse en el resto de base, resto de azúcar o esqueleto de fosfato para mejorar, por ejemplo, la estabilidad, hibridación o solubilidad de la molécula. Por ejemplo, el esqueleto de desoxirribosa-fosfato de los ácidos nucleicos puede modificarse para generar ácidos nucleicos peptídicos (véase Hyrup *et al.*, 1996, *Bioorganic & Medicinal Chemistry* 4(1):5-23). Tal como se usan en el presente documento, los términos “ácidos nucleicos peptídicos” o “PNA” (por sus siglas en inglés) se refieren a miméticos de ácido nucleico, por ejemplo, mimético de ADN, en los que el esqueleto de desoxirribosa-fosfato se reemplaza por un esqueleto pseudopeptídico y sólo se retienen las cuatro nucleobases naturales. Se ha demostrado que el esqueleto neutro de los PNA permite la hibridación específica a ADN y ARN en condiciones de baja fuerza iónica. La síntesis de oligómeros de PNA puede realizarse usando protocolos de síntesis de péptidos en fase sólida convencionales tal como se describe en Hyrup *et al.* (1996), citado anteriormente; Perry-O’ Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

Los PNA pueden usarse en aplicaciones terapéuticas y de diagnóstico. Por ejemplo, los PNA pueden usarse como agentes antisentido o antígenos para la modulación específica de secuencia de la expresión génica mediante, por ejemplo, inducción de la transcripción o traducción deteniendo o inhibiendo la replicación. Los PNA también pueden usarse, por ejemplo, en el análisis de mutaciones de un único par de bases en un gen por, por ejemplo, bloqueo por PCR dirigida por PNA; como enzimas de restricción artificial cuando se usan en combinación con otras enzimas, por ejemplo, nucleasas S1 (Hyrup (1996), citado anteriormente; o como sondas o cebadores para la secuencia de ADN y la hibridación (Hyrup, 1996, citado anteriormente; Perry-O’ Keefe *et al.*, 1996, *Proc. Natl. Acad. Sci. USA* 93:14670-675).

En otra realización, los PNA pueden modificarse, por ejemplo, para potenciar su estabilidad o captación celular, uniendo grupos lipófilos u otros grupos auxiliares a PNA, mediante la formación de quimeras de PNA-ADN, o mediante el uso de liposomas u otras técnicas de administración de fármacos conocidas en la técnica. Por ejemplo, pueden generarse quimeras de ADN-PNA que pueden combinar las propiedades ventajosas de PNA y ADN. Tales quimeras permiten que enzimas de reconocimiento de ADN, por ejemplo, ARNasa H y ADN polimerasas, interaccionen con la parte de ADN mientras que la parte de PNA proporcionará una alta afinidad y especificidad de unión. Las quimeras de ADN-PNA pueden ligarse usando ligadores de longitudes apropiadas seleccionadas en términos de apilamiento de bases, número de enlaces entre las nucleobases y orientación (Hyrup, 1996, citado anteriormente). La síntesis de quimeras de ADN-PNA puede realizarse tal como se describe en Hyrup (1996), citado anteriormente, y Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63. Por ejemplo, una cadena de ADN puede sintetizarse sobre un soporte sólido usando química de acoplamiento de fosforamidita convencional y análogos de nucleósidos modificados. Los compuestos tales como 5’-(4-metoxitritil)amino-5’-desoxi-timidina-fosforamidita pueden usarse como una unión entre el PNA y el extremo 5’ del ADN (Mag *et al.*, 1989, *Nucleic Acids Res.* 17:5973-88). Luego, los monómeros de PNA se acoplan de forma escalonada para producir una molécula quimérica con un segmento de PNA en 5’ y un segmento de ADN en 3’ (Finn *et al.*, 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternativamente, pueden sintetizarse moléculas quiméricas con un segmento de ADN en 5’ y un segmento de PNA en 3’ (Peterser *et al.*, 1975, *Bioorganic Med. Chem Lett.* 5:1119-11124).

En otras realizaciones, el oligonucleótido puede incluir otros grupos colgantes tales como péptidos (por ejemplo, para seleccionar como diana receptores de la célula huésped *in vivo*), o agentes que facilitan el transporte a través de la membrana celular (véanse, por ejemplo, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:648-652; la publicación PCT n.º WO 88/09810) o la barrera hematoencefálica (véase, por ejemplo, la publicación PCT n.º WO 89/10134). Además, los oligonucleótidos pueden modificarse con agentes de escisión activados por hibridación (véase, por ejemplo, Krol *et al.*, 1988, *Bio/Techniques* 6:958-976) o agentes intercalantes (véase, por ejemplo, Zon, 1988, *Pharm. Res.* 5:539-549). Con este propósito, el oligonucleótido puede conjugarse con otra molécula, por ejemplo, un péptido, un agente de reticulación activado por hibridación, un agente de transporte, un agente de escisión activado por hibridación, etc.

La invención también incluye ácidos nucleicos de baliza molecular que tienen al menos una región que es complementaria a un ácido nucleico de la invención, de tal manera que la baliza molecular es útil para cuantificar la presencia del ácido nucleico de la invención en una muestra. Un ácido nucleico de “baliza molecular” es un ácido nucleico que comprende un par de regiones complementarias y que tiene un fluoróforo y un extintor fluorescente asociado con el mismo. El fluoróforo y el extintor están asociados con diferentes partes del ácido nucleico en una orientación tal que cuando las regiones complementarias se aparean entre sí, el extintor extingue la fluorescencia del fluoróforo. Cuando las regiones complementarias del ácido nucleico no se hibridan entre sí, la fluorescencia del fluoróforo se extingue en menor grado. Los ácidos nucleicos de baliza molecular se describen, por ejemplo, en la patente estadounidense 5.876.930.

D. Proteínas aisladas y anticuerpos

Un aspecto de la invención se refiere a proteínas marcadoras aisladas y partes biológicamente activas de las

mismas, así como a fragmentos de polipéptido adecuados para su uso como inmunógenos para generar anticuerpos dirigidos contra una proteína marcadora o un fragmento de la misma. En una realización, la proteína marcadora nativa puede aislarse de fuentes de tejido o células mediante un esquema de purificación apropiado usando técnicas de purificación de proteína convencionales. En otra realización, una proteína o un péptido que comprende la totalidad o un segmento de la proteína marcadora se produce mediante técnicas de ADN recombinante. De manera alternativa a la expresión recombinante, tal proteína o péptido puede sintetizarse químicamente usando técnicas de síntesis de péptidos convencionales.

Una proteína "aislada" o "purificada" o una parte biológicamente activa de la misma está sustancialmente libre de material celular u otras proteínas contaminantes de la fuente celular o tisular de la que deriva la proteína, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente. La expresión "sustancialmente libre de material celular" incluye preparaciones de proteína en las que la proteína se separa de los componentes celulares de las células de las que se aísla o se produce de manera recombinante. Por tanto, la proteína que está sustancialmente libre de material celular incluye preparaciones de proteína que tienen menos de aproximadamente el 30%, el 20%, el 10% o el 5% (en peso seco) de proteína heteróloga (también denominada en el presente documento "proteína contaminante"). Cuando la proteína o parte biológicamente activa de la misma se produce de manera recombinante, también está preferiblemente sustancialmente libre de medio de cultivo, es decir, el medio de cultivo representa menos del 20%, el 10% o el 5% del volumen de la preparación de proteína. Cuando la proteína se produce mediante síntesis química, preferiblemente está sustancialmente libre de precursores químicos u otros productos químicos, es decir, se separa de los precursores químicos u otros productos químicos que están implicados en la síntesis de la proteína. Por consiguiente, tales preparaciones de la proteína tienen menos de aproximadamente el 30%, el 20%, el 10%, el 5% (en peso seco) de precursores químicos o compuestos distintos del polipéptido de interés.

Las partes biológicamente activas de una proteína marcadora incluyen polipéptidos que comprenden secuencias de aminoácidos suficientemente idénticas o derivadas de la secuencia de aminoácidos de la proteína marcadora, que incluyen menos aminoácidos que la proteína de longitud completa, y presentan al menos una actividad de la proteína de longitud completa correspondiente. Normalmente, las partes biológicamente activas comprenden un dominio o motivo con al menos una actividad de la proteína de longitud completa correspondiente. Una parte biológicamente activa de una proteína marcadora de la invención puede ser un polipéptido que tiene, por ejemplo, 10, 25, 50, 100 o más aminoácidos de longitud. Además, otras partes biológicamente activas, en las que se delecionan otras regiones de la proteína marcadora, pueden prepararse mediante técnicas recombinantes y evaluarse para una o más de las actividades funcionales de la forma nativa de la proteína marcadora.

Las proteínas marcadoras preferidas están codificadas por secuencias de nucleótidos que comprenden las secuencias que codifican para cualquiera de los genes descritos en los ejemplos. Otras proteínas útiles son sustancialmente idénticas (por ejemplo, al menos aproximadamente el 40%, preferiblemente el 50%, el 60%, el 70%, el 80%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99%) a una de estas secuencias y retienen la actividad funcional de la proteína marcadora que se produce de manera natural correspondiente, aunque difieren en la secuencia de aminoácidos debido a mutagénesis o la variación alélica natural.

Para determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos ácidos nucleicos, las secuencias se alinean con propósitos de comparación óptima (por ejemplo, pueden introducirse huecos en la secuencia de una primera secuencia de aminoácidos o de ácido nucleico para una alineación óptima con una segunda secuencia de aminoácidos o de ácido nucleico). Luego se comparan los residuos de aminoácido o nucleótidos en las posiciones de aminoácido o posiciones de nucleótido correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. Preferiblemente, el porcentaje de identidad entre las dos secuencias se calcula usando una alineación global. Alternativamente, el porcentaje de identidad entre las dos secuencias se calcula usando una alineación local. El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % de identidad = # de posiciones idénticas/# total de posiciones (por ejemplo, posiciones solapantes) x100). En una realización, las dos secuencias tienen la misma longitud. En otra realización, las dos secuencias no tienen la misma longitud.

La determinación del porcentaje de identidad entre dos secuencias puede lograrse usando un algoritmo matemático. Un ejemplo preferido y no limitativo de un algoritmo matemático usado para la comparación de dos secuencias es el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-2268, modificado como en Karlin y Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5877. Dicho algoritmo se incorpora en los programas BLASTN y BLASTX de Altschul *et al.* (1990) J. Mol. Biol. 215:403-410. Las búsquedas de nucleótidos en BLAST pueden realizarse con el programa BLASTN, puntuación = 100, longitud de palabra = 12 para obtener secuencias de nucleótidos homólogas a las moléculas de ácido nucleico de la invención. Las búsquedas de proteína en BLAST pueden realizarse con el programa BLASTP, puntuación = 50, longitud de palabra = 3 para obtener secuencias de aminoácidos homólogas a las moléculas de proteína de la invención. Para obtener alineaciones con huecos con propósitos de comparación, puede usarse una versión más nueva del algoritmo BLAST denominado Gapped BLAST tal como se describe

en Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402, que puede realizar alineaciones locales con huecos para los programas BLASTN, BLASTP y BLASTX. Alternativamente, PSI-Blast puede usarse para realizar una búsqueda iterativa que detecta relaciones distantes entre moléculas. Cuando se utilizan los programas BLAST, Gapped BLAST y PSI-Blast, pueden usarse los parámetros por defecto de los programas respectivos (por ejemplo, BLASTX y BLASTN). Véase <http://www.ncbi.nlm.nih.gov>. Otro ejemplo preferido y no limitativo de un algoritmo matemático usado para la comparación de secuencias es el algoritmo de Myers y Miller, (1988) *CABIOS* 4:11-17. Dicho algoritmo se incorpora en el programa ALIGN (versión 2.0) que forma parte del paquete de software de alineación de secuencias GCG. Cuando se usa el programa ALIGN para comparar secuencias de aminoácidos, puede usarse una tabla de residuos de peso PAM120, una penalización por longitud de hueco de 12 y una penalización por hueco de 4. Otro algoritmo útil para identificar regiones de similitud de secuencia local y alineación es el algoritmo FASTA tal como se describe en Pearson y Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444-2448. Cuando se usa el algoritmo FASTA para comparar secuencias de nucleótidos o de aminoácidos puede usarse, por ejemplo, una tabla de residuos de peso PAM120 con un valor k-tuple de 2.

El porcentaje de identidad entre dos secuencias puede determinarse usando técnicas similares a las descritas anteriormente, con o sin permitir huecos. Al calcular el porcentaje de identidad, sólo se cuentan las coincidencias exactas.

La invención también proporciona proteínas quiméricas o de fusión que comprenden una proteína marcadora o un segmento de la misma. Tal como se usa en el presente documento, una "proteína quimérica" o "proteína de fusión" comprende la totalidad o parte (preferiblemente una parte biológicamente activa) de una proteína marcadora operativamente unida a un polipéptido heterólogo (es decir, un polipéptido distinto de la proteína marcadora). Dentro de la proteína de fusión, el término "operativamente unido" pretende indicar que la proteína marcadora o el segmento de la misma y el polipéptido heterólogo se fusionan en marco entre sí. El polipéptido heterólogo puede fusionarse con el extremo amino-terminal o el extremo carboxilo-terminal de la proteína marcadora o su segmento.

Una proteína de fusión útil es una proteína de fusión de GST en la que una proteína marcadora o su segmento se fusiona al extremo carboxilo-terminal de las secuencias de GST. Tales proteínas de fusión pueden facilitar la purificación de un polipéptido recombinante de la invención.

En otra realización, la proteína de fusión contiene una secuencia señal heteróloga en su extremo amino-terminal. Por ejemplo, la secuencia señal nativa de una proteína marcadora puede eliminarse y reemplazarse por una secuencia señal de otra proteína. Por ejemplo, la secuencia secretora de gp67 de la proteína de envuelta de baculovirus puede usarse como secuencia señal heteróloga (Ausubel *et al.*, Ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, 1992). Otros ejemplos de secuencias señal heterólogas eucariotas incluyen las secuencias secretoras de melitina y la fosfatasa alcalina placentaria humana (Stratagene; La Jolla, California). En aún otro ejemplo, las secuencias señal heterólogas procariontas útiles incluyen la señal secretora de *phoA* (Sambrook *et al.*, citado anteriormente) y la señal secretora de proteína A (Pharmacia Biotech; Piscataway, Nueva Jersey).

En aún otra realización, la proteína de fusión es una proteína de fusión de inmunoglobulina en la que la totalidad o parte de una proteína marcadora se fusiona a secuencias derivadas de un miembro de la familia de proteínas de inmunoglobulina. Las proteínas de fusión de inmunoglobulina de la invención pueden incorporarse en composiciones farmacéuticas y administrarse a un sujeto para inhibir una interacción entre un ligando (soluble o unido a la membrana) y una proteína en la superficie de una célula (receptor), para suprimir de ese modo la transducción de señales *in vivo*. La proteína de fusión de inmunoglobulina puede usarse para afectar a la biodisponibilidad de un ligando relacionado de una proteína marcadora. La inhibición de la interacción ligando/receptor puede ser útil terapéuticamente, tanto para tratar trastornos proliferativos y diferenciativos como para modular (por ejemplo, promover o inhibir) la supervivencia celular. Además, las proteínas de fusión de inmunoglobulina de la invención pueden usarse como inmunógenos para producir anticuerpos dirigidos contra una proteína marcadora en un sujeto, para purificar ligandos y en ensayos de examen para identificar moléculas que inhiben la interacción de la proteína marcadora con ligandos.

Las proteínas quiméricas y de fusión de la invención pueden producirse mediante técnicas de ADN recombinante convencionales. En otra realización, el gen de fusión puede sintetizarse mediante técnicas convencionales que incluyen sintetizadores de ADN automatizados. Alternativamente, la amplificación por PCR de fragmentos de genes puede llevarse a cabo usando cebadores de anclaje que dan lugar a proyecciones complementarias entre dos fragmentos de genes consecutivos que posteriormente pueden aparearse y volverse a amplificar para generar una secuencia génica quimérica (véase, por ejemplo, Ausubel *et al.*, citado anteriormente). Además, muchos vectores de expresión están disponibles comercialmente que ya codifican para un resto de fusión (por ejemplo, un polipéptido de GST). Un ácido nucleico que codifica para un polipéptido de la invención puede clonarse en tal vector de expresión de tal manera que el resto de fusión esté ligado en marco al polipéptido de la invención.

Puede usarse una secuencia señal para facilitar la secreción y el aislamiento de proteínas marcadoras. Las secuencias señal se caracterizan normalmente por un núcleo de aminoácidos hidrófobos que generalmente se escinden de la proteína madura durante la secreción en uno o más eventos de escisión. Tales péptidos señal

5 contienen sitios de procesamiento que permiten la escisión de la secuencia señal de las proteínas maduras a medida que pasan a través de la ruta secretora. Por tanto, la invención se refiere a proteínas marcadoras, proteínas de fusión o segmentos de las mismas que tienen una secuencia señal, así como a tales proteínas de las que la secuencia señal se ha escindido proteolíticamente (es decir, los productos de escisión). En una realización, una secuencia de ácido nucleico que codifica para una secuencia señal puede ligarse operativamente en un vector de expresión a una proteína de interés, tal como una proteína marcadora o un segmento de la misma. La secuencia señal dirige la secreción de la proteína, tal como desde un huésped eucariota en el que se transforma el vector de expresión, y la secuencia señal se escinde posteriormente o simultáneamente. La proteína puede purificarse fácilmente del medio extracelular mediante métodos reconocidos en la técnica. Alternativamente, la secuencia señal puede ligarse a la proteína de interés usando una secuencia que facilite la purificación, como con un dominio GST.

15 La presente invención también se refiere a variantes de las proteínas marcadoras. Tales variantes tienen una secuencia de aminoácidos alterada que pueden funcionar o bien como agonistas (miméticos) o bien como antagonistas. Las variantes pueden generarse mediante mutagénesis, por ejemplo, mutación puntual discreta o truncamiento. Un agonista puede retener sustancialmente la misma, o un subconjunto, de las actividades biológicas de la forma que se produce de manera natural de la proteína. Un antagonista de una proteína puede inhibir una o más de las actividades de la forma que se produce de manera natural de la proteína, por ejemplo, uniéndose de manera competitiva a un miembro posterior o anterior de una cascada de señalización celular que incluye la proteína de interés. Por tanto, el tratamiento con una variante de función limitada puede provocar efectos biológicos específicos. El tratamiento de un sujeto con una variante que tiene un subconjunto de las actividades biológicas de la forma que se produce de manera natural de la proteína puede tener menos efectos secundarios en un sujeto en relación con el tratamiento con la forma que se produce de manera natural de la proteína.

25 Las variantes de una proteína marcadora que funcionan o bien como agonistas (miméticos) o bien como antagonistas pueden identificarse mediante el examen de bibliotecas combinatorias de mutantes, por ejemplo, mutantes de truncamiento, de la proteína de la invención para determinar actividad agonista o antagonista. En una realización, se genera una biblioteca variada de variantes mediante mutagénesis combinatoria a nivel de ácido nucleico y está codificada por una biblioteca genética variada. Puede producirse una biblioteca variada de variantes, por ejemplo, ligando enzimáticamente una mezcla de oligonucleótidos sintéticos en secuencias génicas de modo que un grupo degenerado de secuencias de proteínas potenciales pueda expresarse como polipéptidos individuales, o alternativamente, como un grupo de proteína de fusión más grandes (por ejemplo, para presentación en fago). Hay una variedad de métodos que pueden usarse para producir bibliotecas de variantes potenciales de las proteínas marcadoras a partir de una secuencia de oligonucleótidos degenerados. Los métodos para sintetizar oligonucleótidos degenerados se conocen en la técnica (véanse, por ejemplo, Narang, 1983, Tetrahedron 39:3; Itakura *et al.*, 1984, Annu. Rev. Biochem. 53:323; Itakura *et al.*, 1984, Science 198:1056; Ike *et al.*, 1983 Nucleic Acid Res. 11:477).

40 Además, las bibliotecas de segmentos de una proteína marcadora pueden usarse para generar una población variada de polipéptidos para el examen y la posterior selección de proteínas marcadoras variantes o segmentos de las mismas. Por ejemplo, puede generarse una biblioteca de fragmentos de secuencia codificante tratando un fragmento de PCR bicatenario de la secuencia codificante de interés con una nucleasa en condiciones en las que se produce mella sólo una vez por molécula, desnaturalizando el ADN bicatenario, renaturalizando el ADN para formar ADN bicatenario que puede incluir pares sentido/antisentido de diferentes productos mellados, retirando partes monocatenarias de dúplex reformados mediante tratamiento con nucleasa S1 y ligando la biblioteca de fragmentos resultante en un vector de expresión. Mediante este método, puede derivar una biblioteca de expresión que codifica para fragmentos amino-terminales e internos de diversos tamaños de la proteína de interés.

50 Se conocen en la técnica diversas técnicas para examinar productos genéticos de bibliotecas combinatorias producidas mediante mutaciones puntuales o truncamiento, y para examinar bibliotecas de ADNc para determinar productos genéticos que tienen una propiedad seleccionada. Las técnicas más usadas, que son susceptibles de un análisis de alto rendimiento, para examinar grandes bibliotecas de genes incluyen normalmente la clonación de la biblioteca de genes en vectores de expresión replicables, la transformación de células apropiadas con la biblioteca de vectores resultante y la expresión de los genes combinatorios en condiciones en las que la detección de una actividad deseada facilita el aislamiento del vector que codifica para el gen cuyo producto se detectó. La mutagénesis de conjunto recursivo (REM), una técnica que potencia la frecuencia de mutantes funcionales en las bibliotecas, puede usarse en combinación con los ensayos de examen para identificar variantes de una proteína de la invención (Arkin y Yourvan, 1992, Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave *et al.*, 1993, Protein Engineering 6 (3):327-331).

60 Otro aspecto de la invención se refiere a anticuerpos dirigidos contra una proteína de la invención. En realizaciones preferidas, los anticuerpos se unen específicamente a una proteína marcadora o un fragmento de la misma. Los términos "anticuerpo" y "anticuerpos", tal como se usan indistintamente en el presente documento, se refieren a moléculas de inmunoglobulina, así como a fragmentos y derivados de las mismas que comprenden una parte inmunológicamente activa de una molécula de inmunoglobulina, (es decir, tal parte contiene un sitio de unión a antígeno que se une específicamente a un antígeno, tal como una proteína marcadora, por ejemplo, un epítipo de una proteína marcadora). Un anticuerpo que se une específicamente a una proteína de la invención es un anticuerpo

que se une a la proteína, pero no se une sustancialmente a otras moléculas en una muestra, por ejemplo, una muestra biológica, que contiene de manera natural la proteína. Los ejemplos de una parte inmunológicamente activa de una molécula de inmunoglobulina incluyen, pero no se limitan a, anticuerpos de cadena sencilla (Accs), fragmentos F(ab) y F(ab')₂.

5 Una proteína aislada de la invención o un fragmento de la misma puede usarse como inmunógeno para generar anticuerpos. Puede usarse la proteína de longitud completa o, como alternativa, la invención proporciona fragmentos de péptidos antigénicos para su uso como inmunógenos. El péptido antigénico de una proteína de la invención comprende al menos 8 (preferiblemente 10, 15, 20 ó 30 o más) residuos de aminoácido de la secuencia de aminoácidos de una de las proteínas de la invención, y abarca al menos un epítipo de la proteína de tal manera que un anticuerpo generado contra el péptido forma un complejo inmunitario específico con la proteína. Los epítopos preferidos abarcados por el péptido antigénico son regiones que están ubicadas en la superficie de la proteína, por ejemplo, regiones hidrófilas. Pueden usarse análisis de secuencia de hidrofobicidad, análisis de secuencia de hidrofiliidad o análisis similares para identificar regiones hidrófilas. En realizaciones preferidas, una proteína marcadora aislada o fragmento de la misma se usa como inmunógeno.

Un inmunógeno se usa normalmente para preparar anticuerpos inmunizando un sujeto adecuado (es decir inmunocompetente) como un conejo, una cabra, un ratón u otro mamífero o vertebrado. Una preparación inmunogénica apropiada puede contener, por ejemplo, proteína o péptido expresado de manera recombinante o sintetizado químicamente. La preparación puede incluir además un adyuvante, tales como adyuvante completo o incompleto de Freund, o un agente inmunoestimulador similar. Composiciones inmunógenas preferidas son aquellas que no contienen otras proteínas humanas tales como, por ejemplo, composiciones inmunógenas producidas usando una célula huésped no humana para la expresión recombinante de una proteína de la invención. De tal manera, las composiciones de anticuerpos resultantes tienen una unión reducida o nula de proteínas humanas distintas de una proteína de la invención.

La invención proporciona anticuerpos policlonales y monoclonales. El término "anticuerpo monoclonal" o "composición de anticuerpo monoclonal", tal como se usa en el presente documento, se refiere a una población de moléculas de anticuerpo que contienen sólo una especie de un sitio de unión a antígeno capaz de inmunorreaccionar con un epítipo particular. Composiciones de anticuerpos policlonales y monoclonales preferidas son aquellas que se han seleccionado para anticuerpos dirigidos contra una proteína de la invención. Preparaciones de anticuerpos policlonales y monoclonales particularmente preferidas son aquellas que contienen sólo anticuerpos dirigidos contra una proteína marcadora o un fragmento de la misma.

Los anticuerpos policlonales pueden prepararse inmunizando a un sujeto adecuado con una proteína de la invención como inmunógeno. El título de anticuerpos en el sujeto inmunizado puede monitorizarse a lo largo del tiempo mediante técnicas convencionales, tal como con un ensayo de inmunosorción ligada a enzimas (ELISA) usando polipéptido inmovilizado. En un momento apropiado después de la inmunización, por ejemplo, cuando los títulos de anticuerpos específicos son los más altos, pueden obtenerse células productoras de anticuerpos del sujeto y usarse para preparar anticuerpos monoclonales (AcM) mediante técnicas convencionales, tales como la técnica del hibridoma descrita originariamente por Kohler y Milstein (1975) *Nature* 256:495-497, la técnica del hibridoma de células B humanas (véase Kozbor *et al.*, 1983, *Immunol. Today* 4:72), la técnica del hibridoma de VEB (véase Cole *et al.*, págs. 77-96 en *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985) o técnicas de trioma. La tecnología para producir hibridomas se conoce bien (véase en general *Current Protocols in Immunology*, Coligan *et al.* ed., John Wiley & Sons, Nueva York, 1994). Las células de hibridoma que producen un anticuerpo monoclonal de la invención se detectan examinando los sobrenadantes de cultivo de hibridoma en busca de anticuerpos que se unan al polipéptido de interés, por ejemplo, usando un ensayo ELISA convencional.

De manera alternativa a la preparación de hibridomas secretores de anticuerpos monoclonales, un anticuerpo monoclonal dirigido contra una proteína de la invención puede identificarse y aislarse examinando una biblioteca de inmunoglobulina combinatoria recombinante (por ejemplo, una biblioteca de presentación en fago de anticuerpos) con el polipéptido de interés. Los kits para generar y examinar bibliotecas de presentación en fago están disponibles comercialmente (por ejemplo, el sistema de anticuerpos en fago recombinante de Pharmacia, n.º de catálogo 27-9400-01; y el kit de presentación en fago *SurfZAP* de Stratagene, n.º de catálogo 240612). Además, pueden hallarse ejemplos de métodos y reactivos particularmente susceptibles de uso en la generación y el examen de la biblioteca de presentación de anticuerpos, por ejemplo, en la patente estadounidense n.º 5.223.409; la publicación PCT n.º WO 92/18619; la publicación PCT n.º WO 91/17271; la publicación PCT n.º WO 92/20791; la publicación PCT n.º WO 92/15679; la publicación PCT n.º WO 93/01288; la publicación PCT n.º WO 92/01047; la publicación PCT n.º WO 92/09690; la publicación PCT n.º WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734.

La invención también proporciona anticuerpos recombinantes que se unen específicamente a una proteína de la invención. En realizaciones preferidas, los anticuerpos recombinantes se unen específicamente a una proteína marcadora o un fragmento de la misma. Los anticuerpos recombinantes incluyen, pero no se limitan a, anticuerpos monoclonales quiméricos y humanizados, que comprenden partes tanto humanas como no humanas, anticuerpos de

cadena sencilla y anticuerpos multiespecíficos. Un anticuerpo quimérico es una molécula en la que diferentes partes derivan de diferentes especies animales, tales como aquellos que tienen una región variable derivada de un AcM murino y una región constante de inmunoglobulina humana. (Véanse, por ejemplo, Cabilly *et al.*, la patente estadounidense n.º 4.816,567; y Boss *et al.*, la patente estadounidense n.º 4.816,397.). Los anticuerpos de cadena sencilla tienen un sitio de unión a antígeno y consisten en un único polipéptido. Pueden producirse mediante técnicas conocidas en la técnica, por ejemplo, usando los métodos descritos en Ladner *et al.* la patente estadounidense n.º 4.946.778; Bird *et al.*, (1988) *Science* 242:423-426; Whitlow *et al.*, (1991) *Methods in Enzymology* 2:1-9; Whitlow *et al.*, (1991) *Methods in Enzymology* 2:97-105; y Huston *et al.*, (1991) *Methods in Enzymology Molecular Design and Modeling: Concepts and Applications* 203:46-88. Los anticuerpos multiespecíficos son moléculas de anticuerpo que tienen al menos dos sitios de unión a antígeno que se unen específicamente a diferentes antígenos. Tales moléculas pueden producirse mediante técnicas conocidas en la técnica, por ejemplo, usando los métodos descritos en Segal, la patente estadounidense n.º 4.676.980; Holliger *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:6444-6448; Whitlow *et al.*, (1994) *Protein Eng.* 7:1017-1026 y la patente estadounidense n.º 6.121.424.

Los anticuerpos humanizados son moléculas de anticuerpos de especies no humanas que tienen una o más regiones determinantes de complementariedad (CDR) de las especies no humanas y una región de entramado de una molécula de inmunoglobulina humana. (Véase, por ejemplo, Queen, la patente estadounidense n.º 5.585.089.). Los anticuerpos monoclonales humanizados pueden producirse mediante técnicas de ADN recombinante conocidas en la técnica, por ejemplo, usando los métodos descritos en la publicación PCT n.º WO 87/02671; la solicitud de patente europea 184.187; la solicitud de patente europea 171.496; la solicitud de patente europea 173.494; la publicación PCT n.º WO 86/01533; la patente estadounidense n.º 4.816.567; la solicitud de patente europea 125.023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; y Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Bio/Techniques* 4:214; la patente estadounidense 5.225.539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoevan *et al.* (1988) *Science* 239:1534; y Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

Más particularmente, pueden producirse anticuerpos humanizados, por ejemplo, usando ratones transgénicos que son incapaces de expresar genes de cadenas pesadas y ligeras de inmunoglobulina endógenos, pero que pueden expresar genes de cadenas pesadas y ligeras humanos. Los ratones transgénicos se inmunizan de manera normal con un antígeno seleccionado, por ejemplo, la totalidad o una parte de un polipéptido correspondiente a un marcador de la invención. Los anticuerpos monoclonales dirigidos contra el antígeno pueden obtenerse usando tecnología del hibridoma convencional. Los transgenes de inmunoglobulina humana que albergan los ratones transgénicos se reorganizan durante la diferenciación de células B, y posteriormente experimentan cambio de clase y mutación somática. Por tanto, usando tal técnica, es posible producir anticuerpos IgG, IgA e IgE terapéuticamente útiles. Para obtener una descripción general de esta tecnología para producir anticuerpos humanos, véase Lonberg y Huszar (1995) *Int. Rev. Immunol.* 13:65-93). Para un análisis detallado de esta tecnología para producir anticuerpos humanos y anticuerpos monoclonales humanos y protocolos para producir tales anticuerpos, véanse, por ejemplo, la patente estadounidense 5.625.126; la patente estadounidense 5.633.425; la patente estadounidense 5.569.825; la patente estadounidense 5.661.016; y la patente estadounidense 5.545.806. Además, empresas como Abgenix, Inc. (Freemont, CA), pueden comprometerse a proporcionar anticuerpos humanos dirigidos contra un antígeno seleccionado usando tecnología similar a la descrita anteriormente.

Los anticuerpos completamente humanos que reconocen un epítipo seleccionado pueden generarse usando una técnica denominada "selección guiada". En este enfoque, un anticuerpo monoclonal no humano seleccionado, por ejemplo, un anticuerpo murino, se usa para guiar la selección de un anticuerpo completamente humano que reconoce el mismo epítipo (Jespers *et al.*, 1994, *Bio/technology* 12:899-903).

Los anticuerpos de la invención pueden aislarse después de la producción (por ejemplo, de la sangre o el suero del sujeto) o síntesis y purificarse adicionalmente mediante técnicas bien conocidas. Por ejemplo, los anticuerpos IgG pueden purificarse usando cromatografía de proteína A. Pueden seleccionarse anticuerpos específicos para una proteína de la invención o (por ejemplo, purificarse parcialmente) o purificarse mediante, por ejemplo, cromatografía de afinidad. Por ejemplo, una proteína de la invención expresada de manera recombinante y purificada (o purificada parcialmente) se produce tal como se describe en el presente documento, y se acopla de manera covalente o no covalente a un soporte sólido tal como, por ejemplo, una columna de cromatografía. La columna puede usarse entonces para purificar por afinidad anticuerpos específicos para las proteínas de la invención a partir de una muestra que contiene anticuerpos dirigidos contra un gran número de epítopos diferentes, generándose de ese modo una composición de anticuerpos sustancialmente purificada, es decir, una que esté sustancialmente libre de anticuerpos contaminantes. Por una composición de anticuerpos sustancialmente purificada se entiende, en este contexto, que la muestra de anticuerpo contiene como máximo sólo el 30% (en peso seco) de anticuerpos contaminantes dirigidos contra epítopos distintos de los de la proteína deseada de la invención, y preferiblemente como máximo el 20%, aún más preferiblemente como máximo el 10%, y lo más preferiblemente como máximo el 5% (en peso seco) de la muestra es anticuerpos contaminantes. Una composición de anticuerpos purificada significa que al menos el 99% de los anticuerpos en la composición están dirigidos contra la proteína deseada de la

invención.

En una realización preferida, los anticuerpos sustancialmente purificados de la invención pueden unirse específicamente a un péptido señal, una secuencia secretada, un dominio extracelular, un dominio transmembrana o citoplasmático o una membrana citoplasmática de una proteína de la invención. En una realización particularmente preferida, los anticuerpos sustancialmente purificados de la invención se unen específicamente a una secuencia secretada o un dominio extracelular de las secuencias de aminoácidos de una proteína de la invención. En una realización más preferida, los anticuerpos sustancialmente purificados de la invención se unen específicamente a una secuencia secretada o un dominio extracelular de las secuencias de aminoácidos de una proteína marcadora.

Un anticuerpo dirigido contra una proteína de la invención puede usarse para aislar la proteína mediante técnicas convencionales, tales como cromatografía de afinidad o inmunoprecipitación. Además, tal anticuerpo puede usarse para detectar la proteína marcadora o el fragmento de la misma (por ejemplo, en un lisado celular o sobrenadante celular) para evaluar el nivel y el patrón de expresión del marcador. Los anticuerpos también pueden usarse en diagnóstico para monitorizar los niveles de proteína en los tejidos o líquidos corporales (por ejemplo, en líquido corporal asociado a estado de enfermedad o estado de toxicidad) como parte de un procedimiento de pruebas clínicas, por ejemplo para, por ejemplo, determinar la eficacia de un régimen de tratamiento dado. La detección puede facilitarse mediante el uso de un derivado de anticuerpo, que comprende un anticuerpo de la invención acoplado a una sustancia detectable. Los ejemplos de sustancias detectables incluyen diversas enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes y materiales radiactivos. Los ejemplos de enzimas adecuadas incluyen peroxidasa de rábano picante, fosfatasa alcalina, β -galactosidasa o acetilcolinesterasa; los ejemplos de complejos de grupos prostéticos adecuados incluyen estreptavidina/biotina y avidina/biotina; los ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamina-fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de un material luminiscente incluye luminol; los ejemplos de materiales bioluminiscentes incluyen luciferasa, luciferina y aequorina, y los ejemplos de material radiactivo adecuado incluyen ^{125}I , ^{131}I , ^{35}S o ^3H .

Los anticuerpos de la invención también pueden usarse como agentes terapéuticos en el tratamiento del cáncer. En una realización preferida, los anticuerpos completamente humanos de la invención se usan para el tratamiento terapéutico de pacientes con cáncer humanos, particularmente aquellos que tienen un cáncer. En otra realización preferida, los anticuerpos que se unen específicamente a una proteína marcadora o un fragmento de la misma se usan para el tratamiento terapéutico. Además, tal anticuerpo terapéutico puede ser un derivado de anticuerpo o inmunotoxina que comprende un anticuerpo conjugado con un resto terapéutico tal como una citotoxina, un agente terapéutico o un ion metálico radiactivo. Una citotoxina o un agente citotóxico incluye cualquier agente que sea perjudicial para las células. Los ejemplos incluyen taxol, citocalasina B, gramicidina D, bromuro de etidio, emetina, mitomicina, etopósido, tenopósido, vincristina, vinblastina, colchicina, doxorubicina, daunorubicina, dihidroxiantracindiona, mitoxantrona, mitramicina, actinomicina D, 1-deshidrotestosterona, glucocorticoides, procaína, tetracaína, lidocaína, propranolol y puomicina y análogos u homólogos de los mismos. Los agentes terapéuticos incluyen, entre otros, antimetabolitos (por ejemplo, metotrexato, 6-mercaptopurina, 6-tioguanina, citarabina, 5-fluorouracilo-decarbazona), agentes alquilantes (por ejemplo, mecloretamina, tioepa-clorambucilo, melfalán, carmustina (BSNU) y lomustina (CCNU), ciclofosfamida, busulfano, dibromomanitol, estreptozotocina, mitomicina C y cis-diclorodiamina-platino (II) (DDP) cisplatino), antraciclinas (por ejemplo, daunorubicina (anteriormente daunomicina) y doxorubicina), antibióticos (por ejemplo, dactinomicina (anteriormente actinomicina), bleomicina, mitramicina y antramina (AMC)), y agentes antimetabólicos (por ejemplo, vincristina y vinblastina).

Los anticuerpos conjugados de la invención pueden usarse para modificar una respuesta biológica dada, ya que el resto de fármaco no debe interpretarse como limitado a los agentes terapéuticos químicos clásicos. Por ejemplo, el resto de fármaco puede ser una proteína o un polipéptido que presenta una actividad biológica deseada. Tales proteínas pueden incluir, por ejemplo, una toxina tal como proteína inhibidora de ribosomas (véase Better *et al.*, la patente estadounidense n.º 6.146.631), abrina, ricina A, exotoxina de *Pseudomonas* o toxina diftérica; una proteína tal como factor de necrosis tumoral, α -interferón, β -interferón, factor de crecimiento nervioso, factor de crecimiento derivado de plaquetas, activador de plasminógeno tisular; o modificadores de la respuesta biológica tales como, por ejemplo, linfocinas, interleucina-1 ("IL-1"), interleucina-2 ("IL-2"), interleucina-6 ("IL-6"), factor estimulante de colonias de granulocitos y macrófagos ("GM-CSF"), factor estimulante de colonias de granulocitos ("G-CSF") u otros factores de crecimiento.

Las técnicas para conjugar tal resto terapéutico con anticuerpos se conocen bien, véanse, por ejemplo, Arnon *et al.*, "Monoclonal Antibodies for Immunotargeting of Drugs in Cancer Therapy", en *Monoclonal Antibodies and Cancer Therapy*, Reisfeld *et al.* (eds.), págs. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", en *Controlled Drug Delivery* (2ª Ed.), Robinson *et al.* (eds.), págs. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents in Cancer Therapy: A Review", en *Monoclonal Antibodies' 84: Biological and Clinical Applications*, Pinchera *et al.* (eds.), págs. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", en *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), págs. 303-16 (Academic Press 1985) y Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58 (1982).

Por consiguiente, en un aspecto, la invención proporciona anticuerpos, fragmentos de anticuerpo y derivados sustancialmente purificados, todos los cuales se unen específicamente a una proteína de la invención y preferiblemente, una proteína marcadora. En diversas realizaciones, los anticuerpos sustancialmente purificados de la invención, o fragmentos o derivados de los mismos, pueden ser anticuerpos humanos, no humanos, quiméricos y/o humanizados. En otro aspecto, la invención proporciona anticuerpos no humanos, fragmentos de anticuerpo y derivados, todos los cuales se unen específicamente a una proteína de la invención y preferiblemente, una proteína marcadora. Tales anticuerpos no humanos pueden ser anticuerpos de cabra, ratón, oveja, caballo, pollo, conejo o rata. Alternativamente, los anticuerpos no humanos de la invención pueden ser anticuerpos quiméricos y/o humanizados. Además, los anticuerpos no humanos de la invención pueden ser anticuerpos policlonales o anticuerpos monoclonales. En otro aspecto adicional, la invención proporciona anticuerpos monoclonales, fragmentos de anticuerpo y derivados, todos los cuales se unen específicamente a una proteína de la invención y preferiblemente, una proteína marcadora. Los anticuerpos monoclonales pueden ser anticuerpos humanos, humanizados, quiméricos y/o no humanos.

La invención también proporciona un kit que contiene un anticuerpo de la invención conjugado con una sustancia detectable, e instrucciones de uso. Todavía otro aspecto de la invención es una composición farmacéutica que comprende un anticuerpo de la invención. En una realización, la composición farmacéutica comprende un anticuerpo de la invención y un portador farmacéuticamente aceptable.

E. Medicina predictiva

La presente invención se refiere al campo de la medicina predictiva en la que se usan ensayos de diagnóstico, ensayos de pronóstico, farmacogenómica y monitorización de ensayos clínicos con propósitos de pronóstico (predictivos) para tratar de ese modo a un individuo de manera profiláctica. Por consiguiente, un aspecto de la presente invención se refiere a ensayos de diagnóstico para determinar el nivel de expresión de una o más proteínas marcadoras o ácidos nucleicos, con el fin de determinar si un individuo está en riesgo de desarrollar determinada enfermedad o toxicidad inducida por fármacos. Tales ensayos pueden usarse con propósitos de pronóstico o predictivos para tratar así de manera profiláctica a un individuo antes del inicio del trastorno.

Aún otro aspecto de la invención se refiere a la monitorización de la influencia de agentes (por ejemplo, fármacos u otros compuestos administrados o bien para inhibir o bien para tratar o bien para prevenir un trastorno o una toxicidad inducida por fármacos {es decir, para comprender cualquier efecto tóxico inducido por fármacos que tal tratamiento pueda tener}) sobre la expresión o actividad de un marcador de la invención en ensayos clínicos. Estos y otros agentes se describen con más detalle en las siguientes secciones.

F. Ensayos de diagnóstico

Un método a modo de ejemplo para detectar la presencia o ausencia de una proteína o un ácido nucleico marcador en una muestra biológica implica obtener una muestra biológica (por ejemplo, muestra de tejido o líquido corporal asociado a toxicidad) de un sujeto de prueba y poner en contacto la muestra biológica con un compuesto o un agente capaz de detectar el polipéptido o el ácido nucleico (por ejemplo, ARNm, ADN genómico o ADNc). Por tanto, los métodos de detección de la invención pueden usarse para detectar ARNm, proteína, ADNc o ADN genómico, por ejemplo, en una muestra biológica *in vitro* así como *in vivo*. Por ejemplo, las técnicas *in vitro* para la detección de ARNm incluyen hibridaciones de tipo Northern e hibridaciones *in situ*. Las técnicas *in vitro* para la detección de una proteína marcadora incluyen ensayos de inmunosorción ligada a enzimas (ELISA), inmunotransferencias de tipo Western, inmunoprecipitaciones e inmunofluorescencia. Las técnicas *in vitro* para la detección de ADN genómico incluyen hibridaciones de tipo Southern. Las técnicas *in vivo* para la detección de ARNm incluyen la reacción en cadena de la polimerasa (PCR), hibridaciones de tipo Northern e hibridaciones *in situ*. Además, las técnicas *in vivo* para la detección de una proteína marcadora incluyen la introducción en un sujeto de un anticuerpo etiquetado dirigido contra la proteína o el fragmento de la misma. Por ejemplo, el anticuerpo puede etiquetarse con un marcador radiactivo cuya presencia y ubicación en un sujeto pueden detectarse mediante técnicas de obtención de imágenes convencionales.

Un principio general de tales ensayos de diagnóstico y pronóstico implica la preparación de una muestra o mezcla de reacción que puede contener un marcador y una sonda, en condiciones apropiadas y durante un tiempo suficiente para permitir que el marcador y la sonda interaccionen y se unan, formando así un complejo que puede retirarse y/o detectarse en la mezcla de reacción. Estos ensayos pueden realizarse de diversas maneras.

Por ejemplo, un método para realizar tal ensayo implicaría anclar el marcador o la sonda sobre un soporte de fase sólida, también denominado sustrato, y detectar complejos de marcador/sonda diana anclados en la fase sólida al final de la reacción. En una realización de tal método, una muestra de un sujeto, que se someterá a ensayo para determinar la presencia y/o concentración del marcador, puede anclarse sobre un portador o soporte de fase sólida. En otra realización, es posible la situación inversa, en la que la sonda puede anclarse a una fase sólida y puede permitirse que una muestra de un sujeto reaccione como componente no anclado del ensayo.

Existen muchos métodos establecidos para anclar los componentes del ensayo a una fase sólida. Estos incluyen, sin

limitación, moléculas marcadoras o de sonda que se inmovilizan a través de la conjugación de biotina y estreptavidina. Tales componentes de ensayo biotinilados pueden prepararse a partir de biotina-NHS (N-hidroxisuccinimida) usando técnicas conocidas en la técnica (por ejemplo, kit de biotinilación, Pierce Chemicals, Rockford, IL) e inmovilizarse en los pocillos de placas de 96 pocillos recubiertas con estreptavidina (Pierce Chemical). En determinadas realizaciones, las superficies con componentes de ensayo inmovilizados pueden prepararse de antemano y almacenarse.

Otros portadores o soportes de fase sólida adecuados para tales ensayos incluyen cualquier material capaz de unirse a la clase de molécula a la que pertenece el marcador o la sonda. Los soportes o portadores bien conocidos incluyen, entre otros, vidrio, poliestireno, nailon, polipropileno, nailon, polietileno, dextrano, amilasas, celulosas naturales y modificadas, poliácridamidas, gabros y magnetita.

Para realizar ensayos con los enfoques mencionados anteriormente, el componente no inmovilizado se añade a la fase sólida sobre la que se ancla el segundo componente. Una vez que se completa la reacción, pueden retirarse los componentes no complejados (por ejemplo, mediante lavado) en condiciones tales que cualquier complejo formado permanecerá inmovilizado en la fase sólida. La detección de complejos marcador/sonda anclados a la fase sólida puede lograrse en varios métodos descritos en el presente documento.

En una realización preferida, la sonda, cuando es el componente de ensayo no anclado, puede etiquetarse con propósitos de detección y lectura del ensayo, directa o indirectamente, con etiquetas detectables comentadas en el presente documento y que conoce bien un experto en la técnica.

También es posible detectar directamente la formación del complejo marcador/sonda sin la manipulación o etiquetado adicional de ninguno de los componentes (marcador o sonda), por ejemplo, usando la técnica de transferencia de energía de fluorescencia (véase, por ejemplo, Lakowicz *et al.*, la patente estadounidense n.º 5.631.169; Stavrianopoulos, *et al.*, la patente estadounidense n.º 4.868.103). Se selecciona una etiqueta de fluoróforo en la primera molécula 'donadora' de tal manera que, tras la excitación con luz incidente de longitud de onda apropiada, su energía fluorescente emitida será absorbida por una etiqueta fluorescente en una segunda molécula 'aceptora', que a su vez es capaz de fluorescer debido a la energía absorbida. Alternativamente, la molécula de proteína 'donadora' puede utilizar simplemente la energía fluorescente natural de los residuos de triptófano. Se eligen etiquetas que emiten a diferentes longitudes de onda de la luz, de modo que la etiqueta de la molécula 'aceptora' pueda diferenciarse de la de la 'donadora'. Dado que la eficiencia de la transferencia de energía entre las etiquetas está relacionada con la distancia que separa las moléculas, pueden evaluarse las relaciones espaciales entre las moléculas. En una situación en la que se produce la unión entre las moléculas, la emisión fluorescente de la etiqueta de la molécula 'aceptora' en el ensayo debe ser máxima. Un evento de unión FET puede medirse convenientemente a través de medios de detección fluorométricos convencionales bien conocidos en la técnica (por ejemplo, usando un fluorímetro).

En otra realización, la determinación de la capacidad de una sonda para reconocer un marcador puede lograrse sin etiquetar ninguno de los componentes del ensayo (sonda o marcador) utilizando una tecnología tal como análisis de interacción biomolecular en tiempo real (BIA) (véanse, por ejemplo, Sjolander, S. y Urbaniczky, C., 1991, *Anal. Chem* 63:2338-2345 y Szabo *et al.*, 1995, *Curr. Opin. Struct. Biol.* 5:699-705). Tal como se usa en el presente documento, "BIA" o "resonancia de plasmón superficial" es una tecnología para estudiar interacciones bioespecíficas en tiempo real, sin etiquetar ninguno de los interaccionantes (por ejemplo, BIAcore). Los cambios en la masa en la superficie de unión (indicativo de un evento de unión) dan como resultado alteraciones del índice de refracción de la luz cerca de la superficie (el fenómeno óptico de resonancia de plasmón superficial (SPR)), lo que da como resultado una señal detectable que puede usarse como indicación de reacciones en tiempo real entre moléculas biológicas.

Alternativamente, en otra realización, pueden realizarse ensayos de diagnóstico y pronóstico análogos con marcador y sonda como solutos en una fase líquida. En tal ensayo, el marcador y la sonda complejados se separan de los componentes no complejados mediante un cualquiera de varias técnicas convencionales, incluyendo pero sin limitarse a: centrifugación diferencial, cromatografía, electroforesis e inmunoprecipitación. En la centrifugación diferencial, los complejos marcador/sonda pueden separarse de los componentes de ensayo no complejados a través de una serie de etapas centrífugas, debido a los diferentes equilibrios de sedimentación de los complejos en función de sus diferentes tamaños y densidades (véase, por ejemplo, Rivas, G. y Minton, A.P., 1993, *Trends Biochem Sci.* 18 (8):284-7). Las técnicas cromatográficas convencionales también pueden usarse para separar las moléculas complejadas de las no complejadas. Por ejemplo, la cromatografía de filtración en gel separa las moléculas en función del tamaño, y mediante la utilización de una resina de filtración en gel apropiada en un formato de columna, por ejemplo, el complejo relativamente más grande puede separarse de los componentes no complejados relativamente más pequeños. De manera similar, las propiedades de carga relativamente diferentes del complejo marcador/sonda en comparación con los componentes no complejados pueden aprovecharse para diferenciar el complejo de los componentes no complejados, por ejemplo mediante la utilización de resinas de cromatografía de intercambio iónico. Tales resinas y técnicas cromatográficas las conoce bien un experto en la técnica (véanse, por ejemplo, Heegaard, N.H., 1998, *J. Mol. Recognit.* Winter 11(1-6):141-8; Hage, D.S. y Tweed, S.A.J *Chromatogr B Biomed Sci Appl* 10 de octubre de 1997; 699(1-2):499-525). La electroforesis en gel también

puede emplearse para separar componentes de ensayo complejados de componentes no unidos (véase, por ejemplo, Ausubel *et al.*, Ed., Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York, 1987-1999). En esta técnica, los complejos de proteína o ácido nucleico se separan según el tamaño o la carga, por ejemplo. Con el fin de mantener la interacción de unión durante el proceso electroforético, normalmente se prefieren materiales y condiciones de la matriz de gel no desnaturizantes en ausencia de agente reductor. Las condiciones apropiadas para el ensayo particular y los componentes del mismo los conocerán bien un experto en la técnica.

En una realización particular, el nivel de ARNm marcador puede determinarse mediante formatos tanto *in situ* como *in vitro* en una muestra biológica usando métodos conocidos en la técnica. El término "muestra biológica" pretende incluir tejidos, células, líquidos biológicos y aislados de los mismos, que se aíslan de un sujeto, así como tejidos, células y líquidos presentes dentro de un sujeto. Muchos métodos de detección de expresión usan ARN aislado. Para los métodos *in vitro*, puede usarse cualquier técnica de aislamiento de ARN que no seleccione en contra del aislamiento de ARNm para la purificación de ARN de las células (véase, por ejemplo, Ausubel *et al.*, Ed., Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York 1987-1999). Además, pueden procesarse fácilmente grandes cantidades de muestras de tejido usando técnicas que conocen bien los expertos en la técnica tales como, por ejemplo, el proceso de aislamiento de ARN de una única etapa de Chomczynski (1989, la patente estadounidense n.º 4.843.155).

El ARNm aislado puede usarse en ensayos de hibridación o amplificación que incluyen, pero no se limitan a, análisis de tipo Southern o Northern, análisis mediante reacción en cadena de la polimerasa y alineamientos de sondas. Un método de diagnóstico preferido para la detección de niveles de ARNm implica poner en contacto el ARNm aislado con una molécula de ácido nucleico (sonda) que puede hibridarse con el ARNm codificado por el gen que está detectándose. La sonda de ácido nucleico puede ser, por ejemplo, un ADNc de longitud completa, o una parte del mismo, tal como un oligonucleótido de al menos 7, 15, 30, 50, 100, 250 ó 500 nucleótidos de longitud y suficiente para hibridar específicamente en condiciones rigurosas a un ARNm o ADN genómico que codifica para un marcador de la presente invención. En el presente documento se describen otras sondas adecuadas para su uso en los ensayos de diagnóstico de la invención. La hibridación de un ARNm con la sonda indica que el marcador en cuestión está expresándose.

En un formato, el ARNm se inmoviliza sobre una superficie sólida y se pone en contacto con una sonda, por ejemplo, haciendo correr el ARNm aislado en un gel de agarosa y transfiriendo el ARNm del gel a una membrana, tal como de nitrocelulosa. En un formato alternativo, la(s) sonda(s) se inmovilizan sobre una superficie sólida y el ARNm se pone en contacto con la(s) sonda(s), por ejemplo, en un alineamiento de chips de genes Affymetrix. Un experto en la técnica puede adaptar fácilmente los métodos de detección de ARNm conocidos para su uso en la detección del nivel de ARNm codificado por los marcadores de la presente invención.

Un método alternativo para determinar el nivel de marcador de ARNm en una muestra implica el proceso de amplificación de ácidos nucleicos, por ejemplo, mediante RT-PCR (la realización experimental expuesta en Mullis, 1987, la patente estadounidense n.º 4.683.202), reacción en cadena de la ligasa (Barany, 1991, Proc. Natl. Acad. Sci. USA, 88:189-193), replicación de secuencia autosostenida (Guatelli *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), sistema de amplificación transcripcional (Kwoh *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), replicasa Q-beta (Lizardi *et al.*, 1988, Bio/Technology 6:1197), replicación por círculo rodante (Lizardi *et al.*, la patente estadounidense n.º 5.854.033) o cualquier otro método de amplificación de ácidos nucleicos, seguido de la detección de las moléculas amplificadas usando técnicas que conocen bien los expertos en la técnica. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas están presentes en cantidades muy bajas. Tal como se usa en el presente documento, los cebadores de amplificación se definen como un par de moléculas de ácido nucleico que pueden hibridarse en regiones en 5' o 3' de un gen (hebras positiva y negativa, respectivamente, o viceversa) y contienen una región corta entremedias. En general, los cebadores de amplificación tienen una longitud de desde aproximadamente 10 hasta 30 nucleótidos y flanquean a una región de desde aproximadamente 50 hasta 200 nucleótidos de longitud. En condiciones apropiadas y con reactivos apropiados, tales cebadores permiten la amplificación de una molécula de ácido nucleico que comprende la secuencia de nucleótidos flanqueada por los cebadores.

Para los métodos *in situ*, no es necesario aislar el ARNm antes de la detección. En tales métodos, se prepara/procesa una muestra celular o tisular usando métodos histológicos conocidos. La muestra se inmoviliza entonces sobre un soporte, normalmente un portaobjetos de vidrio, y luego se pone en contacto con una sonda que puede hibridarse con ARNm que codifica para el marcador.

Como alternativa a la realización de determinaciones basándose en el nivel de expresión absoluto del marcador, las determinaciones pueden basarse en el nivel de expresión normalizado del marcador. Los niveles de expresión se normalizan corrigiendo el nivel de expresión absoluto de un marcador al comparar su expresión con la expresión de un gen que no es un marcador, por ejemplo, un gen de mantenimiento que se expresa de manera constitutiva. Los genes adecuados para la normalización incluyen genes de mantenimiento tales como el gen de la actina o genes específicos de células epiteliales. Esta normalización permite la comparación del nivel de expresión en una muestra, por ejemplo, una muestra de paciente, con otra muestra, por ejemplo, una muestra no tóxica o no enferma, o entre muestras de diferentes fuentes.

Alternativamente, el nivel de expresión puede proporcionarse como un nivel de expresión relativo. Para determinar un nivel de expresión relativo de un marcador, el nivel de expresión del marcador se determina para 10 muestras o más de aislados de células normales frente a de enfermedad o tóxicos, preferiblemente 50 muestras o más, antes de la determinación del nivel de expresión para la muestra en cuestión. Se determina el nivel de expresión medio de cada uno de los genes sometidos a ensayo en el mayor número de muestras y se usa esto como nivel de expresión de referencia para el marcador. El nivel de expresión del marcador determinado para la muestra de prueba (nivel absoluto de expresión) se divide entonces entre el valor de expresión medio obtenido para ese marcador. Esto proporciona un nivel de expresión relativo.

Preferiblemente, las muestras usadas en la determinación de referencia serán de células no tóxicas o no enfermas. La elección de la fuente de la célula depende del uso del nivel de expresión relativo. El uso de la expresión que se encuentra en los tejidos normales como puntuación de expresión media ayuda a validar si el marcador sometido a ensayo es específico de enfermedad o toxicidad (frente a células normales). Además, a medida que se acumulan más datos, puede revisarse el valor de expresión medio, proporcionando valores de expresión relativos mejorados basándose en los datos acumulados. Los datos de expresión de células de enfermedad o células tóxicas proporcionan un medio para graduar la gravedad de la enfermedad o el estado tóxico.

En otra realización de la presente invención, se detecta una proteína marcadora. Un agente preferido para detectar la proteína marcadora de la invención es un anticuerpo capaz de unirse a tal proteína o a un fragmento de la misma, preferiblemente un anticuerpo con una etiqueta detectable. Los anticuerpos pueden ser policlonales, o más preferiblemente, monoclonales. Puede usarse un anticuerpo intacto, o un fragmento o derivado del mismo (por ejemplo, Fab o F(ab')₂). El término "etiquetado", con respecto a la sonda o el anticuerpo, pretende abarcar el etiquetado directo de la sonda o el anticuerpo mediante acoplamiento (es decir, ligamiento físico) de una sustancia detectable a la sonda o el anticuerpo, así como etiquetado indirecto de la sonda o el anticuerpo por reactividad con otro reactivo que se etiqueta directamente. Los ejemplos de etiquetado indirecto incluyen la detección de un anticuerpo primario usando un anticuerpo secundario etiquetado con fluorescencia y el etiquetado de extremos de una sonda de ADN con biotina de modo que pueda detectarse con estreptavidina etiquetada con fluorescencia.

Pueden aislarse proteínas de las células usando técnicas que conocen bien los expertos en la técnica. Los métodos de aislamiento de proteínas empleados pueden ser, por ejemplo, los descritos en Harlow y Lane (Harlow y Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York).

Puede emplearse una variedad de formatos para determinar si una muestra contiene una proteína que se une a un anticuerpo dado. Los ejemplos de tales formatos incluyen, pero no se limitan a, inmunoensayo enzimático (EIA), radioinmunoensayo (RIA), análisis de inmunotransferencia de tipo Western y ensayo de inmunoabsorción ligada a enzimas (ELISA). Un experto en la técnica puede adaptar fácilmente métodos conocidos de detección de proteínas/anticuerpos para su uso en la determinación de si las células expresan un marcador de la presente invención.

En un formato, pueden usarse anticuerpos, o fragmentos o derivados de anticuerpo, en métodos tales como inmunotransferencias de tipo Western o técnicas de inmunofluorescencia para detectar las proteínas expresadas. En tales usos, generalmente es preferible inmovilizar o bien el anticuerpo o bien las proteínas sobre un soporte sólido. Los soportes o portadores de fase sólida adecuados incluyen cualquier soporte capaz de unir un antígeno o un anticuerpo. Los soportes o portadores bien conocidos incluyen vidrio, poliestireno, polipropileno, polietileno, dextrano, nailon, amilasas, celulosas naturales y modificadas, poliácridamidas, gabros y magnetita.

Un experto en la técnica conocerá muchos otros portadores adecuados para la unión de anticuerpo o antígeno, y podrá adaptar tal soporte para su uso con la presente invención. Por ejemplo, la proteína aislada de las células de enfermedad o tóxicas puede hacerse correr en una electroforesis en gel de poliácridamida e inmovilizarse sobre un soporte de fase sólida tal como nitrocelulosa. El soporte puede lavarse entonces con tampones adecuados seguido de tratamiento con el anticuerpo etiquetado de manera detectable. El soporte de fase sólida puede lavarse entonces con el tampón por segunda vez para eliminar el anticuerpo no unido. La cantidad de etiqueta unida en el soporte sólido puede detectarse mediante medios convencionales.

La invención también abarca kits para detectar la presencia de una proteína o un ácido nucleico marcador en una muestra biológica. Tales kits pueden usarse para determinar si un sujeto padece o corre un mayor riesgo de desarrollar determinadas enfermedades o toxicidad inducida por fármacos. Por ejemplo, el kit puede comprender un compuesto o agente etiquetado capaz de detectar una proteína o un ácido nucleico marcador en una muestra biológica y medios para determinar la cantidad de la proteína o el ARNm en la muestra (por ejemplo, un anticuerpo que se une a la proteína o un fragmento de la misma, o una sonda de oligonucleótidos que se une al ADN o ARNm que codifica para la proteína). Los kits también pueden incluir instrucciones para interpretar los resultados obtenidos con el kit.

Para los kits basados en anticuerpos, el kit puede comprender, por ejemplo: (1) un primer anticuerpo (por ejemplo, unido a un soporte sólido) que se une a una proteína marcadora; y, opcionalmente, (2) un segundo anticuerpo

diferente que se une o bien a la proteína o bien al primer anticuerpo y se conjuga con una etiqueta detectable.

Para los kits basados en oligonucleótidos, el kit puede comprender, por ejemplo: (1) un oligonucleótido, por ejemplo, un oligonucleótido etiquetado de manera detectable, que se hibrida con una secuencia de ácido nucleico que codifica para una proteína marcadora o (2) un par de cebadores útiles para amplificar una molécula de ácido nucleico marcador. El kit también puede comprender, por ejemplo, un agente tamponante, un conservante o un agente de estabilización de proteínas. El kit puede comprender además componentes necesarios para detectar la etiqueta detectable (por ejemplo, una enzima o un sustrato). El kit también puede contener una muestra de control o una serie de muestras de control que pueden someterse a ensayo y compararse con la muestra de prueba. Cada componente del kit puede incluirse dentro de un recipiente individual y todos los diversos recipientes pueden estar dentro de un único envase, junto con instrucciones para interpretar los resultados de los ensayos realizados usando el kit.

G. Farmacogenómica

Los marcadores de la invención también son útiles como marcadores farmacogenómicos. Tal como se usa en el presente documento, un "marcador farmacogenómico" es un marcador bioquímico diana cuyo nivel de expresión se correlaciona con una susceptibilidad o respuesta al fármaco clínica específica en un paciente (véase, por ejemplo, McLeod *et al.* (1999) *Eur. J. Cancer* 35 (12):1650-1652). La presencia o cantidad de la expresión de marcadores farmacogenómicos está relacionada con la respuesta predicha del paciente y más particularmente de las células enfermas o tóxicas del paciente a la terapia con un fármaco o clase de fármacos específicos. Al evaluar la presencia o cantidad de la expresión de uno o más marcadores farmacogenómicos en un paciente, puede seleccionarse una terapia farmacológica que sea la más apropiada para el paciente, o que se predice que tendrá un mayor grado de éxito. Por ejemplo, basándose en la presencia o cantidad de ARN o proteína codificada por marcadores tumorales específicos en un paciente, puede seleccionarse un fármaco o ciclo de tratamiento que esté optimizado para el tratamiento del tumor específico que probablemente esté presente en el paciente. Por tanto, el uso de marcadores farmacogenómicos permite seleccionar o diseñar el tratamiento más apropiado para cada paciente con cáncer sin probar diferentes fármacos o regímenes.

Otro aspecto de la farmacogenómica trata sobre estados genéticos que alteran la manera en que el cuerpo actúa sobre los fármacos. Estos estados farmacogenéticos pueden aparecer o bien como defectos raros o bien como polimorfismos. Por ejemplo, la deficiencia de glucosa-6-fosfato deshidrogenasa (G6PD) es una enzimopatía hereditaria común en la que la principal complicación clínica es la hemólisis después de la ingestión de fármacos oxidantes (antipalúdicos, sulfonamidas, analgésicos, nitrofuranos) y consumo de habas.

Como una realización ilustrativa, la actividad de las enzimas metabolizadoras de fármacos es un determinante importante tanto de la intensidad como de la duración de la acción del fármaco. El descubrimiento de polimorfismos genéticos de enzimas metabolizadoras de fármacos (por ejemplo, N-acetiltransferasa 2 (NAT 2) y las enzimas del citocromo P450, CYP2D6 y CYP2C19) han explicado por qué algunos pacientes no obtienen los efectos farmacológicos esperados o muestran una respuesta exagerada al fármaco y una toxicidad grave después de tomar la dosis convencional y segura de un fármaco. Estos polimorfismos se expresan en dos fenotipos en la población, el metabolizador rápido (EM, *extensive metabolizer*) y el metabolizador lento (PM, *poor metabolizer*). La prevalencia de PM es diferente entre las diferentes poblaciones. Por ejemplo, el gen que codifica para CYP2D6 es altamente polimórfico y se han identificado diversas mutaciones en PM, que conducen todas a la ausencia de CYP2D6 funcional. Los metabolizadores lentos de CYP2D6 y CYP2C19 con bastante frecuencia experimentan una respuesta exagerada al fármaco y efectos secundarios cuando reciben dosis convencionales. Si un metabolito es el resto terapéutico activo, un PM no mostrará respuesta terapéutica, tal como se demuestra por el efecto analgésico de la codeína mediada por su metabolito formado por CYP2D6, la morfina. El otro extremo son los denominados metabolizadores ultrarrápidos que no responden a las dosis convencionales. Recientemente, se ha identificado que la base molecular del metabolismo ultrarrápido se debe a la amplificación del gen de CYP2D6.

Por tanto, puede determinarse el nivel de expresión de un marcador de la invención en un individuo para seleccionar de ese modo el/los agente(s) apropiado(s) para el tratamiento terapéutico o profiláctico del individuo. Además, pueden usarse estudios farmacogenéticos para aplicar genotipos de alelos polimórficos que codifican para enzimas metabolizadoras de fármacos a la identificación del fenotipo de respuesta a fármacos de un individuo. Este conocimiento, cuando se aplica a la dosificación o selección de fármacos, puede evitar reacciones adversas o fracaso terapéutico y, por tanto, potenciar la eficacia terapéutica o profiláctica cuando se trata a un sujeto con un modulador de expresión de un marcador de la invención.

H. Monitorización de ensayos clínicos

La monitorización de la influencia de los agentes (por ejemplo, compuestos farmacológicos) en el nivel de expresión de un marcador de la invención puede aplicarse no sólo en el examen básico de fármacos, sino también en ensayos clínicos. Por ejemplo, la efectividad de un agente para afectar a la expresión de marcadores puede monitorizarse en ensayos clínicos de sujetos que reciben tratamiento para determinadas enfermedades, tales como cáncer, diabetes, obesidad, enfermedad cardiovascular y cardiotoxicidad, o toxicidad inducida por fármacos. En una realización

preferida, la presente invención proporciona un método para monitorizar la efectividad del tratamiento de un sujeto con un agente (por ejemplo, un agonista, antagonista, peptidomimético, proteína, péptido, ácido nucleico, molécula pequeña u otro candidato a fármaco) que comprende las etapas de (i) obtener una muestra previa a la administración de un sujeto antes de la administración del agente; (ii) detectar el nivel de expresión de uno o más marcadores seleccionados de la invención en la muestra previa a la administración; (iii) obtener una o más muestras posteriores a la administración del sujeto; (iv) detectar el nivel de expresión del/de los marcador(es) en las muestras posteriores a la administración; (v) comparar el nivel de expresión del/de los marcador(es) en la muestra o muestras posteriores a la administración con el nivel de expresión del/de los marcador(es) en la muestra o muestras posteriores a la administración; y (vi) alterar la administración del agente al sujeto en consecuencia. Por ejemplo, una mayor expresión del/de los gen(es) marcador(es) durante el transcurso del tratamiento puede indicar una dosificación ineficaz y la conveniencia de aumentar la dosificación. Por el contrario, la disminución de la expresión del/de los gen(es) marcador(es) puede indicar un tratamiento eficaz y no es necesario cambiar la dosificación.

H. Alineamientos

La invención también incluye un alineamiento que comprende un marcador de la presente invención. El alineamiento puede usarse para someter a ensayo la expresión de uno o más genes en el alineamiento. En una realización, el alineamiento puede usarse para someter a ensayo la expresión génica en un tejido para determinar la especificidad tisular de los genes en el alineamiento. De esta manera, pueden someterse a ensayo simultáneamente hasta 7600 genes para determinar la expresión. Esto permite que se desarrolle un perfil que muestre una batería de genes expresados específicamente en uno o más tejidos.

Además de tal determinación cualitativa, la invención permite la cuantificación de la expresión génica. Por tanto, no sólo puede determinarse la especificidad tisular, sino también el nivel de expresión de una batería de genes en el tejido. Por tanto, los genes pueden agruparse basándose en su expresión tisular *per se* y el nivel de expresión en ese tejido. Esto es útil, por ejemplo, para determinar la relación de expresión génica entre unos tejidos y otros. Por tanto, puede perturbarse un tejido y puede determinarse el efecto sobre la expresión génica en un segundo tejido. En este contexto, puede determinarse el efecto de un tipo de células sobre otro tipo de células en respuesta a un estímulo biológico. Tal determinación es útil, por ejemplo, para conocer el efecto de la interacción célula-célula a nivel de la expresión génica. Si un agente se administra terapéuticamente para tratar un tipo de célula pero tiene un efecto indeseable sobre otro tipo de célula, la invención proporciona un ensayo para determinar la base molecular del efecto indeseable y, por tanto, brinda la oportunidad de coadministrar un agente de acción contraria o tratar de otro modo el efecto indeseable. De manera similar, incluso dentro de un tipo de célula individual, pueden determinarse los efectos biológicos indeseables a nivel molecular. Por tanto, los efectos de un agente sobre la expresión de otro gen distinto a la diana pueden determinarse y contrarrestarse.

En otra realización, el alineamiento puede usarse para monitorizar el transcurso de tiempo de expresión de uno o más genes en el alineamiento. Esto puede producirse en diversos contextos biológicos, tal como se describe en el presente documento, por ejemplo, desarrollo de toxicidad inducida por fármacos, progresión de toxicidad inducida por fármacos y procesos, tales como transformación celular asociada con toxicidad inducida por fármacos.

El alineamiento también es útil para determinar el efecto de la expresión de un gen sobre la expresión de otros genes en la misma célula o en células diferentes. Esto proporciona, por ejemplo, una selección de dianas moleculares alternativas para la intervención terapéutica si la diana final o posterior no puede regularse.

El alineamiento también es útil para determinar patrones de expresión diferencial de uno o más genes en células normales y anómalas. Esto proporciona una batería de genes que podrían servir como diana molecular para el diagnóstico o la intervención terapéutica.

VII. Métodos para obtener muestras

Las muestras útiles en los métodos de la invención incluyen cualquier muestra de tejido, célula, biopsia o líquido corporal que exprese un marcador de la invención. En una realización, una muestra puede ser un tejido, una célula, sangre completa, suero, plasma, raspado bucal, saliva, líquido cefalorraquídeo, orina, heces o lavado broncoalveolar. En realizaciones preferidas, la muestra de tejido es una muestra de estado de enfermedad o de estado de toxicidad. En realizaciones más preferidas, la muestra de tejido es una muestra de cáncer, una muestra de diabetes, una muestra de obesidad, una muestra cardiovascular o una muestra de toxicidad inducida por fármacos.

Pueden obtenerse muestras corporales de un sujeto mediante una variedad de técnicas conocidas en la técnica incluyendo, por ejemplo, el uso de una biopsia o raspando o frotando un área o usando una aguja para aspirar líquidos corporales. Los métodos para recoger diversas muestras corporales se conocen bien en la técnica.

Las muestras de tejido adecuadas para detectar y cuantificar un marcador de la invención pueden ser frescas, congeladas o fijadas según métodos conocidos por un experto en la técnica. Las muestras de tejido adecuadas se seccionan preferiblemente y se colocan sobre un portaobjetos de microscopio para análisis adicionales.

Alternativamente, las muestras sólidas, es decir, las muestras de tejido pueden solubilizarse y/u homogeneizarse y posteriormente analizarse como extractos solubles.

5 En una realización, una muestra de biopsia recién obtenida se congela usando, por ejemplo, nitrógeno líquido o difluorodichlorometano. La muestra congelada se prepara para seccionarse usando, por ejemplo, OCT, y se secciona en serie en un criostato. Las secciones en serie se recogen sobre un portaobjetos de microscopio de vidrio. Para la tinción inmunohistoquímica, los portaobjetos pueden recubrirse, por ejemplo, con alumbre de cromo, gelatina o poli-L-lisina para garantizar que las secciones se adhieran a los portaobjetos. En otra realización, las muestras se fijan e incrustan antes de seccionarse. Por ejemplo, una muestra de tejido puede fijarse, por ejemplo, en formalina, deshidratarse en serie e incrustarse, por ejemplo, en parafina.

15 Una vez que se obtiene la muestra, puede usarse cualquier método conocido en la técnica que sea adecuado para detectar y cuantificar un marcador de la invención (o bien a nivel de ácido nucleico o bien de proteína). Tales métodos se conocen bien en la técnica e incluyen, pero no se limitan a, inmunotransferencias de tipo Western, transferencias de tipo Northern, transferencias de tipo Southern, inmunohistoquímica, ELISA, por ejemplo, ELISA amplificado, inmunoprecipitación, inmunofluorescencia, citometría de flujo, inmunocitoquímica, análisis mediante espectrometría de masas, por ejemplo, MALDI-TOF y SELDI-TOF, técnicas de hibridación de ácidos nucleicos, métodos de transcripción inversa de ácidos nucleicos y métodos de amplificación de ácidos nucleicos. En realizaciones particulares, la expresión de un marcador de la invención se detecta a nivel de proteína usando, por ejemplo, anticuerpos que se unen específicamente a estas proteínas.

20 Puede ser necesario tener que modificar las muestras para que un marcador de la invención sea accesible para la unión de anticuerpos. En un aspecto particular de los métodos de inmunocitoquímica o inmunohistoquímica, los portaobjetos pueden transferirse a un tampón de pretratamiento y calentarse opcionalmente para aumentar la accesibilidad de antígeno. El calentamiento de la muestra en el tampón de pretratamiento perturba rápidamente la bicapa lipídica de las células y hace que los antígenos (puede ser el caso en muestras frescas, pero no es normalmente lo que sucede en muestras fijadas) sean más accesibles para la unión de anticuerpos. Los términos "tampón de pretratamiento" y "tampón de preparación" se usan indistintamente en el presente documento para referirse a un tampón que se usa para preparar muestras de citología o histología para inmunotinción, particularmente aumentando la accesibilidad de un marcador de la invención para la unión de anticuerpos. El tampón de pretratamiento puede comprender una solución salina de pH específico, un polímero, un detergente o un tensioactivo no iónico o aniónico tal como, por ejemplo, un tensioactivo aniónico o no iónico etiloxilado, un alcanato o un alcoxilato o incluso mezclas de estos tensioactivos o incluso el uso de una sal biliar. El tampón de pretratamiento puede ser, por ejemplo, una disolución del 0,1% al 1% de sal de sodio de ácido desoxicólico, o una disolución de laureth-13-carboxilato de sodio (por ejemplo, Sandopan LS) o/y complejo aniónico etoxilado. En algunas realizaciones, el tampón de pretratamiento también puede usarse como tampón de almacenamiento de portaobjetos.

40 Cualquier método para hacer que las proteínas marcadoras de la invención sean más accesibles para la unión de anticuerpos puede usarse en la práctica de la invención, incluyendo los métodos de recuperación de antígeno conocidos en la técnica. Véanse, por ejemplo, Bibbo *et al.* (2002) *Acta. Citol.* 46:25-29; Saqi *et al.* (2003) *Diagn. Cytopathol.* 27:365-370; Bibbo *et al.* (2003) *Anal. Quant. Cytol. Histol.* 25:8-11.

45 Tras el pretratamiento para aumentar la accesibilidad de la proteína marcadora, las muestras pueden bloquearse usando un agente de bloqueo apropiado, por ejemplo, un reactivo de bloqueo de peroxidasa tal como peróxido de hidrógeno. En algunas realizaciones, las muestras pueden bloquearse usando un reactivo de bloqueo de proteínas para impedir la unión inespecífica del anticuerpo. El reactivo de bloqueo de proteínas puede comprender, por ejemplo, caseína purificada. Luego se incubaba con la muestra un anticuerpo, particularmente un anticuerpo monoclonal o policlonal que se une específicamente a un marcador de la invención. Un experto en la técnica apreciará que puede obtenerse un pronóstico o diagnóstico más preciso en algunos casos mediante la detección de múltiples epítomos en una proteína marcadora de la invención en una muestra de paciente. Por tanto, en realizaciones particulares, se usan al menos dos anticuerpos dirigidos a diferentes epítomos de un marcador de la invención. Cuando se usa más de un anticuerpo, estos anticuerpos pueden añadirse a una única muestra secuencialmente como reactivos de anticuerpo individuales o simultáneamente como un cóctel de anticuerpos. Alternativamente, cada anticuerpo individual puede añadirse a una muestra independientes del mismo paciente, y agruparse los datos resultantes.

60 Las técnicas para detectar la unión de anticuerpos se conocen bien en la técnica. La unión de anticuerpos a un marcador de la invención puede detectarse mediante el uso de reactivos químicos que generan una señal detectable que corresponde al nivel de unión de anticuerpos y, por consiguiente, al nivel de expresión de la proteína marcadora. En uno de los métodos de inmunohistoquímica o inmunocitoquímica de la invención, la unión de anticuerpos se detecta mediante el uso de un anticuerpo secundario que se conjuga con un polímero etiquetado. Los ejemplos de polímeros etiquetados incluyen, pero no se limitan a, conjugados polímero-enzima. Las enzimas en estos complejos se usan normalmente para catalizar la deposición de un cromógeno en el sitio de unión antígeno-anticuerpo, dando como resultado de ese modo una tinción celular que corresponde al nivel de expresión del biomarcador de interés. Las enzimas de interés particular incluyen, pero no se limitan a, peroxidasa de rábano picante (HRP) y fosfatasa

alcalina (AP).

En un método de inmunohistoquímica o inmunocitoquímica particular de la invención, la unión de anticuerpos a un marcador de la invención se detecta mediante el uso de un polímero etiquetado con HRP que se conjuga con un anticuerpo secundario. La unión de anticuerpos también puede detectarse mediante el uso de un reactivo de sonda específica de especie, que se une a anticuerpos monoclonales o policlonales, y un polímero conjugado con HRP, que se une al reactivo de sonda específica de especie. Los portaobjetos se tiñen para determinar la unión de anticuerpos usando cualquier cromógeno, por ejemplo, el cromógeno 3,3-diaminobencidina (DAB), y luego se someten a contratinción con hematoxilina y, opcionalmente, un agente de azulado tal como hidróxido de amonio o TBS/Tween-20. Otros cromógenos adecuados incluyen, por ejemplo, 3-amino-9-etilcarbazol (AEC). En algunos aspectos de la invención, los portaobjetos se revisan microscópicamente por un técnico de citología y/o un patólogo para evaluar la tinción celular, por ejemplo, tinción fluorescente (es decir, expresión de marcador). Alternativamente, las muestras pueden revisarse a través de microscopía automatizada o por personal con la asistencia de un software informático que facilite la identificación de células con tinción positiva.

La detección de la unión de anticuerpos puede facilitarse mediante el acoplamiento de anticuerpos anti-marcador a una sustancia detectable. Los ejemplos de sustancias detectables incluyen diversas enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes y materiales radiactivos. Los ejemplos de enzimas adecuadas incluyen peroxidasa de rábano picante, fosfatasa alcalina, β -galactosidasa o acetilcolinesterasa; los ejemplos de complejos de grupos prostéticos adecuados incluyen estreptavidina/biotina y avidina/biotina; los ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamina-fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de un material luminiscente incluye luminol; los ejemplos de materiales bioluminiscentes incluyen luciferasa, luciferina y aequorina; y los ejemplos de material radiactivo adecuado incluyen ^{125}I , ^{131}I , ^{35}S , ^{14}C o ^3H .

En una realización de la invención, las muestras congeladas se preparan tal como se describió anteriormente y se tiñen posteriormente con anticuerpos contra un marcador de la invención diluido hasta una concentración apropiada usando, por ejemplo, solución salina tamponada con Tris (TBS). Los anticuerpos primarios pueden detectarse incubando los portaobjetos en anticuerpo anti-inmunoglobulina biotinilado. Esta señal puede amplificarse y visualizarse opcionalmente usando precipitación con diaminobencidina del antígeno. Además, los portaobjetos pueden someterse a contratinción opcionalmente, por ejemplo, con hematoxilina, para visualizar las células.

En otra realización, las muestras fijadas e incrustadas se tiñen con anticuerpos contra un marcador de la invención y se someten a contratinción tal como se describió anteriormente para secciones congeladas. Además, las muestras pueden tratarse opcionalmente con agentes para amplificar la señal con el fin de visualizar la tinción de anticuerpos. Por ejemplo, puede usarse una deposición de biotiniil-tiramida catalizada por peroxidasa, que se hace reaccionar a su vez con estreptavidina conjugada con peroxidasa (sistema de amplificación de señal catalizada (CSA), DAKO, Carpintería, CA).

Los ensayos basados en tejidos (es decir, inmunohistoquímica) son los métodos preferidos para detectar y cuantificar un marcador de la invención. En una realización, la presencia o ausencia de un marcador de la invención puede determinarse por inmunohistoquímica. En una realización, el análisis inmunohistoquímico usa bajas concentraciones de un anticuerpo anti-marcador de manera que las células que carecen del marcador no se tiñen. En otra realización, la presencia o ausencia de un marcador de la invención se determina usando un método inmunohistoquímico que usa altas concentraciones de un anticuerpo anti-marcador de tal manera que las células que carecen de la proteína marcadora se tiñen fuertemente. Las células que no se tiñen o bien contienen marcador mutado y no producen proteína marcadora reconocible antigénicamente, o bien son células en las que las rutas que regulan los niveles de marcador están desreguladas, lo que da como resultado la expresión en estado estacionario de proteína marcadora insignificante.

Un experto en la técnica reconocerá que la concentración de un anticuerpo particular usado para poner en práctica los métodos de la invención variará dependiendo de factores tales como tiempo para la unión, nivel de especificidad del anticuerpo para un marcador de la invención y método de preparación de muestras. Además, cuando se usan múltiples anticuerpos, la concentración requerida puede verse afectada por el orden en que se aplican los anticuerpos a la muestra, por ejemplo, simultáneamente como un cóctel o secuencialmente como reactivos de anticuerpo individuales. Además, la química de detección usada para visualizar la unión de anticuerpos a un marcador de la invención también debe optimizarse para producir la relación señal/ruido deseada.

En una realización de la invención, se usan métodos proteómicos, por ejemplo espectrometría de masas, para detectar y cuantificar las proteínas marcadoras de la invención. Por ejemplo, puede usarse espectrometría de masas de tiempo de vuelo de desorción/ionización láser asociada a matriz (MALDI-TOF EM) o espectrometría de masas de tiempo de vuelo de desorción/ionización láser de superficie mejorada (SELDI-TOF EM) que implica la aplicación de un muestra biológica, tal como suero, a un chip de unión a proteínas (Wright, G.L., Jr., *et al.* (2002) *Expert Rev Mol Diagn* 2:549; Li, J. *et al.* (2002) *Clin Chem* 48:1296; Laronga, C. *et al.* (2003) *Dis Markers* 19:229; Petricoin, E.F., *et al.* (2002) 359:572; Adam, B.L., *et al.* (2002) *Cancer Res* 62:3609; Tolson, J. *et al.* (2004) *Lab Invest* 84:845; Xiao, Z. *et al.* (2001) *Cancer Res* 61:6029) para detectar y cuantificar las proteínas PY-Shc y/o p66-Shc. Se describen

métodos de espectrometría de masas, por ejemplo, en las patentes estadounidenses n.ºs 5.622.824, 5.605.798 y 5.547.835.

5 En otras realizaciones, la expresión de un marcador de la invención se detecta a nivel de ácido nucleico. Las técnicas basadas en ácidos nucleicos para evaluar la expresión se conocen bien en la técnica e incluyen, por ejemplo, determinar el nivel de ARNm marcador en una muestra de un sujeto. Muchos métodos de detección de expresión usan ARN aislado. Cualquier técnica de aislamiento de ARN que no seleccione en contra del aislamiento de ARNm puede usarse para la purificación de ARN de las células que expresan un marcador de la invención (véase, por ejemplo, Ausubel *et al.*, Ed., (1987-1999) *Current Protocols in Molecular Biology* (John Wiley & Sons, Nueva York). Además, pueden procesarse fácilmente grandes cantidades de muestras de tejido usando técnicas que conocen bien los expertos en la técnica tales como, por ejemplo, el proceso de aislamiento de ARN de una única etapa de Chomczynski (1989, la patente estadounidense n.º 4.843.155).

15 El término "sonda" se refiere a cualquier molécula que sea capaz de unirse selectivamente a un marcador de la invención, por ejemplo, un transcrito de nucleótidos y/o proteína. Un experto en la técnica puede sintetizar sondas o derivarlas de preparaciones biológicas apropiadas. Las sondas pueden estar diseñadas específicamente para etiquetarse. Los ejemplos de moléculas que pueden usarse como sondas incluyen, entre otras, ARN, ADN, proteínas, anticuerpos y moléculas orgánicas.

20 El ARNm aislado puede usarse en ensayos de hibridación o amplificación que incluyen, pero no se limitan a, análisis de tipo Southern o Northern, análisis mediante reacción en cadena de polimerasa y alineamientos de sondas. Un método para la detección de los niveles de ARNm implica poner en contacto el ARNm aislado con una molécula de ácido nucleico (sonda) que puede hibridarse con el ARNm marcador. La sonda de ácido nucleico puede ser, por ejemplo, un ADNc de longitud completa, o una parte del mismo, tal como un oligonucleótido de al menos 7, 15, 30, 25 50, 100, 250 ó 500 nucleótidos de longitud y suficiente para hibridarse específicamente en condiciones rigurosas al ADN genómico marcador.

30 En una realización, el ARNm se inmoviliza sobre una superficie sólida y se pone en contacto con una sonda, por ejemplo, haciendo correr el ARNm aislado en un gel de agarosa y transfiriendo el ARNm desde el gel a una membrana, tal como de nitrocelulosa. En una realización alternativa, la(s) sonda(s) se inmovilizan sobre una superficie sólida y el ARNm se pone en contacto con la(s) sonda(s), por ejemplo, en un alineamiento de chips de genes Affymetrix. Un experto en la técnica puede adaptar fácilmente los métodos de detección de ARNm conocidos para su uso en la detección del nivel de ARNm marcador.

35 Un método alternativo para determinar el nivel de ARNm marcador en una muestra implica el proceso de amplificación de ácidos nucleicos, por ejemplo, mediante RT-PCR (la realización experimental expuesta en Mullis, 1987, la patente estadounidense n.º 4.683.202), reacción en cadena de la ligasa (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), replicación de secuencia autosostenida (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), sistema de amplificación transcripcional (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), replicasa Q-beta (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), replicación por círculo rodante (Lizardi *et al.*, la patente estadounidense n.º 5.854.033) o cualquier otro método de amplificación de ácidos nucleicos, seguido de la detección de las moléculas amplificadas usando técnicas que conocen bien los expertos en la técnica. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas están presentes en cantidades muy bajas. En aspectos particulares de la invención, la expresión de 40 marcadores se evalúa mediante RT-PCR fluorogénica cuantitativa (es decir, el sistema TaqMan™). Tales métodos normalmente utilizan pares de cebadores oligonucleotídicos que son específicos para un marcador de la invención. Los métodos para diseñar cebadores oligonucleotídicos específicos para una secuencia conocida se conocen bien en la técnica.

50 Los niveles de expresión de un marcador de la invención pueden monitorizarse usando una transferencia de membrana (tales como la usada en el análisis de hibridación, tal como de tipo Northern, Southern, puntual, y similares), o micropocillos, tubos de muestra, geles, perlas o fibras (o cualquier soporte sólido que comprende ácidos nucleicos unidos). Véanse las patentes estadounidenses n.ºs 5.770.722, 5.874.219, 5.744.305, 5.677.195 y 5.445.934. La detección de la expresión de marcadores también puede comprender el uso de sondas de ácido 55 nucleico en disolución.

60 En una realización de la invención, se usan microalineamientos para detectar la expresión de un marcador de la invención. Los microalineamientos son particularmente adecuados con este propósito debido a la reproducibilidad entre diferentes experimentos. Los microalineamientos de ADN proporcionan un método para la medición simultánea de los niveles de expresión de grandes cantidades de genes. Cada alineamiento consiste en un patrón reproducible de sondas de captura unidas a un soporte sólido. El ARN o ADN etiquetado se hibrida con sondas complementarias en el alineamiento y luego se detecta mediante exploración láser. Las intensidades de hibridación para cada sonda en el alineamiento se determinan y se convierten en un valor cuantitativo que representa los niveles relativos de expresión génica. Véanse las patentes estadounidenses n.ºs 6.040.138, 5.800.992 y 6.020.135, 6.033.860 y 65 6.344.316. Los alineamientos de oligonucleótidos de alta densidad son particularmente útiles para determinar el perfil de expresión génica para un gran número de ARN en una muestra.

Las cantidades de marcador y/o una relación matemática de las cantidades de un marcador de la invención pueden usarse para calcular el riesgo de recidiva de un estado de enfermedad, por ejemplo cáncer, diabetes, obesidad, enfermedad cardiovascular o un estado de toxicidad, por ejemplo, una toxicidad o cardiotoxicidad inducida por fármacos, en un sujeto que está tratándose por un estado de enfermedad o estado de toxicidad, la supervivencia de un sujeto que está tratándose por un estado de enfermedad o un estado de toxicidad, si un estado de enfermedad o estado de toxicidad es agresivo, la eficacia de un régimen de tratamiento para tratar un estado de enfermedad o estado de toxicidad, y similares, usando los métodos de la invención, que pueden incluir métodos de análisis de regresión conocidos por un experto en la técnica. Por ejemplo, los modelos de regresión adecuados incluyen, entre otros, CART (por ejemplo, Hill, T y Lewicki, P. (2006) "STATISTICS Methods and Applications" StatSoft, Tulsa, OK), Cox (por ejemplo, www.evidence-based-medicine.co.uk), exponencial, normal y log normal (por ejemplo, www.obgyn.cam.ac.uk/mrg/statsbook/stsurvan.html), logístico (por ejemplo, www.en.wikipedia.org/wiki/Logistic_regression), paramétrico, no paramétrico, semiparamétrico (por ejemplo, www.socserv.mcmaster.ca/jfox/Books/Companion), lineal (por ejemplo, www.en.wikipedia.org/wiki/Linear_regression), o aditivo (por ejemplo, www.en.wikipedia.org/wiki/Generalized_additive_model).

En una realización, un análisis de regresión incluye las cantidades de marcador. En otra realización, un análisis de regresión incluye una relación matemática de marcador. En aún otra realización, un análisis de regresión de las cantidades de marcador y/o una relación matemática de marcador puede incluir covariables clínicas y/o moleculares adicionales. Tales covariables clínicas incluyen, pero no se limitan a, estado ganglionar, estadio tumoral, grado tumoral, tamaño tumoral, régimen de tratamiento, por ejemplo, quimioterapia y/o radioterapia, desenlace clínico (por ejemplo, recaída, supervivencia específica de la enfermedad, fracaso de la terapia) y/o desenlace clínico en función del tiempo después del diagnóstico, tiempo después del inicio de la terapia y/o tiempo después de completarse el tratamiento.

VIII. Kits

La invención también proporciona composiciones y kits para pronosticar un estado de enfermedad, por ejemplo cáncer, diabetes, obesidad, enfermedad cardiovascular o un estado de toxicidad, por ejemplo, una toxicidad o cardiotoxicidad inducida por fármacos, la recidiva de un estado de enfermedad o estado de toxicidad, o la supervivencia de un sujeto que está tratándose para un estado de enfermedad o estado de toxicidad. Estos kits incluyen uno o más de los siguientes: un anticuerpo detectable que se une específicamente a un marcador de la invención, un anticuerpo detectable que se une específicamente a un marcador de la invención, reactivos para obtener y/o preparar muestras de tejido del sujeto para tinción, e instrucciones de uso.

Los kits de la invención pueden comprender opcionalmente componentes adicionales útiles para realizar los métodos de la invención. A modo de ejemplo, los kits pueden comprender líquidos (por ejemplo, tampón SSC) adecuados para aparear ácidos nucleicos complementarios o para unir un anticuerpo con una proteína con la que se une específicamente, uno o más compartimentos de muestra, un material de instrucciones que describe la realización de un método de la invención y controles/patrones específicos de tejido.

IX. Ensayos de examen

Los dianas de la invención incluyen, pero no se limitan a, los genes y/o las proteínas enumerados en el presente documento. Basándose en los resultados de los experimentos descritos por los solicitantes en el presente documento, las proteínas clave moduladas en un estado de enfermedad o estado de toxicidad están asociadas con o pueden clasificarse en diferentes rutas o grupos de moléculas, incluyendo los componentes del citoesqueleto, los factores de transcripción, respuesta apoptótica, ruta de pentosas fosfato, ruta biosintética, estrés oxidativo (prooxidante), alteraciones de la membrana y metabolismo de fosforilación oxidativa. Por consiguiente, en una realización de la invención, un marcador puede incluir uno o más genes (o proteínas) seleccionados del grupo que consiste en HSPA8, FLNB, PARK7, HSPA1A/HSPA1B, ST13, TUBB3, MIF, KARS, NARS, LGALS1, DDX17, EIF5A, HSPA5, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1, CANX, GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. En una realización, un marcador puede incluir uno o más genes (o proteínas) seleccionados del grupo que consiste en GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1, GPAT1 y TAZ. En algunas realizaciones, los marcadores son una combinación de al menos dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, once, doce, trece, catorce, quince, dieciséis, diecisiete, dieciocho, diecinueve, veinte, veinticinco, treinta o más de los genes (o proteínas) anteriores.

A continuación se describen ensayos de examen útiles para identificar moduladores de marcadores identificados.

La invención también proporciona métodos (también denominados en el presente documento "ensayos de examen") para identificar moduladores, es decir, compuestos o agentes candidatos o de prueba (por ejemplo, proteínas, péptidos, peptidomiméticos, peptoides, moléculas pequeñas u otros fármacos), que son útiles para tratar o prevenir un estado de enfermedad o un estado de toxicidad modulando la expresión y/o actividad de un marcador de la invención. Tales ensayos comprenden normalmente una reacción entre un marcador de la invención y uno o más

componentes de ensayo. Los otros componentes pueden ser o bien el propio compuesto de prueba o bien una combinación de compuestos de prueba y una pareja de unión natural de un marcador de la invención. Los compuestos identificados a través de ensayos tales como los descritos en el presente documento pueden ser útiles, por ejemplo, para modular, por ejemplo, inhibir, mejorar, tratar o prevenir la agresividad de un estado de enfermedad o estado de toxicidad.

Los compuestos de prueba usados en los ensayos de examen de la presente invención pueden obtenerse de cualquier fuente disponible, incluyendo las bibliotecas sistemáticas de compuestos naturales y/o sintéticos. Los compuestos de prueba también pueden obtenerse mediante cualquiera de los numerosos enfoques en los métodos de biblioteca combinatoria conocidos en la técnica, incluyendo: bibliotecas biológicas; bibliotecas de peptoides (bibliotecas de moléculas que tienen las funcionalidades de los péptidos, pero con un nuevo esqueleto no peptídico que son resistentes a la degradación enzimática pero que, sin embargo, siguen siendo bioactivas; véase, por ejemplo, Zuckermann *et al.*, 1994, *J. Med. Chem* 37:2678-85); bibliotecas de fase de disolución o de fase sólida paralelas direccionables espacialmente; métodos de biblioteca sintética que requieren deconvolución; el método de biblioteca "una perla, un compuesto"; y métodos de biblioteca sintética que usan la selección mediante cromatografía de afinidad. Los enfoques de biblioteca biológica y biblioteca de peptoides se limitan a bibliotecas de péptidos, mientras que los otros cuatro enfoques son aplicables a bibliotecas de compuestos de péptidos, oligómeros no peptídicos o moléculas pequeñas (Lam, 1997, *Anticancer Drug Des.* 12:145).

Pueden hallarse ejemplos de métodos para la síntesis de bibliotecas moleculares en la técnica, por ejemplo en: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994) *J. Med. Chem* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem Int. Ed. Engl.* 33:2061; y en Gallop *et al.* (1994) *J. Med. Chem* 37:1233.

Las bibliotecas de compuestos pueden presentarse en disolución (por ejemplo, Houghten, 1992, *Biotechniques* 13:412-421), o en perlas (Lam, 1991, *Nature* 354:82-84), chips (Fodor, 1993, *Nature* 364:555-556), bacterias y/o esporas, (Ladner, documento USP 5.223.409), plásmidos (Cull *et al.*, 1992, *Proc Natl Acad Sci USA* 89:1865-1869) o en fago (Scott y Smith, 1990, *Science* 249:386-390; Devlin, 1990, *Science* 249:404-406; Cwirla *et al.*, 1990, *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici, 1991, *J. Mol. Biol.* 222:301-310; Ladner citado anteriormente.).

Los métodos de examen de la invención comprenden poner en contacto una célula de estado de enfermedad o una célula de estado de toxicidad con un compuesto de prueba y determinar la capacidad del compuesto de prueba para modular la expresión y/o actividad de un marcador de la invención en la célula. La expresión y/o actividad de un marcador de la invención puede determinarse tal como se describe en el presente documento.

En otra realización, la invención proporciona ensayos para examinar compuestos candidatos o de prueba que son sustratos de un marcador de la invención o partes biológicamente activas del mismo. En aún otra realización, la invención proporciona ensayos para examinar compuestos candidatos o de prueba que se unen a un marcador de la invención o partes biológicamente activas del mismo. La determinación de la capacidad del compuesto de prueba para unirse directamente a un marcador puede lograrse, por ejemplo, mediante el acoplamiento del compuesto con un radioisótopo o etiqueta enzimática de tal manera que la unión del compuesto al marcador pueda determinarse mediante la detección del compuesto marcador etiquetado en un complejo. Por ejemplo, pueden etiquetarse compuestos (por ejemplo, sustratos marcadores) con ¹³¹I, ¹²⁵I, ³⁵S, ¹⁴C o ³H, o bien directa o bien indirectamente, y detectarse el radioisótopo mediante recuento directo de radioemisión o por recuento por centelleo. Alternativamente, los componentes del ensayo pueden etiquetarse enzimáticamente, por ejemplo, con peroxidasa de rábano picante, fosfatasa alcalina o luciferasa, y el marcador enzimático detectarse mediante la determinación de la conversión de un sustrato apropiado en producto.

Esta invención se refiere además a nuevos agentes identificados mediante los ensayos de detección descritos anteriormente. Por consiguiente, está dentro del alcance de esta invención usar adicionalmente un agente identificado tal como se describe en el presente documento en un modelo animal apropiado. Por ejemplo, puede usarse un agente capaz de modular la expresión y/o actividad de un marcador de la invención identificado tal como se describe en el presente documento, en un modelo animal para determinar la eficacia, toxicidad o los efectos secundarios del tratamiento con tal agente. Alternativamente, puede usarse un agente identificado tal como se describe en el presente documento, en un modelo animal para determinar el mecanismo de acción de tal agente. Además, esta invención se refiere a usos de nuevos agentes identificados mediante los ensayos de examen descritos anteriormente para el tratamiento tal como se describió anteriormente.

Realizaciones detalladas de la invención

REALIZACIÓN DETALLADA 1: Empleo de la tecnología de plataforma para construir unas redes de simulación u de consenso sobre el cáncer

En esta realización detallada, se empleó la tecnología de plataforma descrita con detalle anteriormente para integrar los datos obtenidos de un modelo de cáncer *in vitro* personalizado, e identificar de ese modo nuevas proteínas/rutas

que controlan la patogenia del cáncer. Los mapas relacionales resultantes de este análisis han proporcionado dianas de tratamiento del cáncer, así como marcadores de diagnóstico/pronóstico asociados con el cáncer.

Se representa el diseño del estudio en la figura 18. En resumen, se sometieron dos líneas celulares de cáncer (PaCa2, HepG2) y una línea celular normal (THLE2) a una de las siete condiciones que simulan un entorno experimentado por las células cancerosas *in vivo*. Específicamente, se expusieron las células a condición hiperglucémica, condición de hipoxia, condición de ácido láctico, condición de combinación hiperglucémica + hipoxia, condición de combinación hiperglucémica + ácido láctico, condición de combinación hipoxia + ácido láctico, o condición de combinación hiperglucémica + hipoxia + ácido láctico. Se crearon diferentes condiciones tales como

- La condición hiperglucémica se creó mediante el cultivo de las células en medios que contenían glucosa 22 mM.
- La condición de hipoxia se indujo al colocar las células en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que se inundó con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno.
- La condición del ácido láctico se creó mediante el cultivo de las células en medios que contenían ácido láctico 12,5 mM.
- La condición de combinación hiperglucémica + hipoxia se creó mediante el cultivo de las células en medios que contenían glucosa 22 mM y se colocaron las células en una cámara de incubador modular inundada con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno.
- La condición de combinación hiperglucémica + ácido láctico se creó mediante el cultivo de las células en medios que contenían glucosa 22 mM y ácido láctico 12,5 mM.
- La condición de combinación hipoxia + ácido láctico se creó mediante el cultivo de las células en medios que contenían ácido láctico 12,5 mM y se colocaron las células en una cámara de incubador modular inundada con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno.
- La condición de combinación hiperglucémica + hipoxia + ácido láctico se creó mediante el cultivo de las células en medios que contenían glucosa 22 mM y ácido láctico 12,5 mM, y se colocaron las células en una cámara de incubador modular inundada con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno.

El modelo celular que comprende las células mencionadas anteriormente, en el que se expusieron las células a cada condición descrita anteriormente, se interrogó adicionalmente exponiendo las células a una perturbación ambiental tratando con coenzima Q10. Específicamente, se trataron las células con coenzima Q10 a 0, 50 μM o 100 μM.

Se recogieron muestras de células, así como muestras de medios para cada línea celular con cada condición y cada tratamiento con coenzima Q10 en diversos momentos después del tratamiento, incluyendo después de 24 horas y 48 horas de tratamiento.

Además, se llevaron a cabo experimentos de interferencia entre dos células cancerosas diferentes, células PaCa2 y células HepG2, en los que se cocultivaron células PaCa2 y HepG2. Este enfoque de cocultivo se denomina experimento de secretoma extracelular (SEC). El primer sistema celular (PaCa2) se sembró en primer lugar en los insertos de los pocillos de una cámara de crecimiento de tipo Transwell. Se usaron placas de seis pocillos para permitir un mejor análisis estadístico. En el momento de la siembra con el primer sistema celular en los insertos, se colocaron los insertos en una placa de 6 pocillos independiente. El segundo sistema celular (HepG2) se sembró en la bandeja primaria. La bandeja de insertos que contenía el primer sistema celular y la bandeja primaria que contenía el segundo sistema celular se incubaron a 37°C durante la noche. Cada uno de los sistemas celulares se hizo crecer en los medios específicos de células específicas (en los que, alternativamente, cada uno de los sistemas celulares podía hacerse crecer en un medio adaptado para soportar el crecimiento de ambos tipos de células). El segundo día, se administró el tratamiento predeterminado mediante intercambio de medios. Específicamente, se colocaron los insertos que contenían el primer sistema celular en la bandeja primaria que contenía el segundo sistema celular. Luego se incubó la bandeja durante un periodo de tiempo predeterminado, por ejemplo, 24 horas o 48 horas. Se configuraron pocillos duplicados con las mismas condiciones, y se agruparon las células para producir material suficiente para el análisis 2D. Los medios (alícuota de 1 ml), las células de los insertos y las células de los pocillos de la bandeja primaria se recogieron como muestras independientes. Los experimentos se realizaron por triplicado para proporcionar un mejor poder de análisis estadístico.

También se llevaron a cabo experimentos de interferencia mediante experimentos de “intercambio de medios”.

Específicamente, se recogió un medio de cultivo o “secretoma” del primer sistema celular (PaCa2) después de 24 horas o 48 horas tras la perturbación o acondicionamiento tal como se describió anteriormente y luego se añadió al segundo sistema celular (HepG2) durante 24-48 horas. Luego se recogió el medio de cultivo final o “secretoma” del segundo sistema celular. Todos los secretomas finales se sometieron a análisis proteómico.

Se realizó iProfiling de los cambios en la expresión de la proteína celular total mediante proteómica cuantitativa para muestras de células y medios recogidas para cada línea celular en cada condición y con cada “perturbación ambiental”, es decir, tratamiento con coenzima Q10, usando las técnicas descritas anteriormente en la descripción detallada. Se realizó de manera similar iProfiling de los cambios en la expresión de proteína celular total mediante proteómica cuantitativa para las muestras de células y medios recogidas para cada línea celular cocultivada en cada condición con cada tratamiento.

Además, se generaron perfiles bioenergéticos de las células cancerosas, células normales y células en experimentos de interferencia expuestas a cada condición y con o sin perturbación de coenzima Q10 empleando el analizador Seahorse esencialmente según lo recomendado por el fabricante. Los electrodos registraron la TCO (tasa de consumo de oxígeno) y TAEC (tasa de acidificación extracelular) en una cámara de 7 μ l creada con el cartucho empujando contra la placa de cultivo de Seahorse.

Se introdujeron todos los datos de proteómica recopilados para cada línea celular (incluyendo células en experimentos de interferencia) en cada condición y con cada perturbación, y los datos de perfiles bioenergéticos recopilados para cada línea celular en cada condición y con cada perturbación, y se procesaron mediante el sistema REFS™. Los datos sin procesar para Paca2, HepG2, THLE2 y experimentos de interferencia se combinaron luego usando una nomenclatura normalizada. Los genes a los que les falta más del 15% de los datos de proteómica se eliminaron por filtrado. Se desarrolló la estrategia de imputación de datos. Por ejemplo, se usó un modelo de error dentro de las réplicas para imputar datos de condiciones experimentales con réplicas. Se usó un algoritmo K-NN basado en 10 vecinos para imputar datos sin réplicas. Se construyeron diferentes modelos de REFS™ para tres sistemas biológicos conjuntamente, sólo para el sistema Paca2 o sólo para el sistema HepG2 vinculado a los datos fenotípicos.

El área bajo la curva y los cambios en veces para cada arista que conecta un nodo primario a un nodo secundario en las redes de simulación los extrajo un programa personalizado que usa el lenguaje de programación R, en el que el lenguaje de programación R es un entorno de software de código abierto para grafos y computación estadística.

Los resultados del programa R se introdujeron en Cytoscape, un programa de código abierto, para generar una representación visual de la red de consenso.

Entre todos los modelos construidos, en la figura 21 se muestra una red de consenso de REFS de interacción proteica a modo de ejemplo a una frecuencia de fragmentos del 70%.

Cada nodo en la red de consenso mostrada en la figura 21 se simuló aumentando o disminuyendo la expresión de LDHA 4 veces para generar una red de simulación usando REFS™, tal como se describió con detalle anteriormente en la descripción detallada.

Se investigó el efecto del cambio de expresión de LDHA simulado sobre PARK7 y proteínas en notas asociadas con PARK7 a alto nivel en la red de consenso a modo de ejemplo mostrada en la figura 21. Se identificaron las proteínas que respondían a la simulación de LDHA en dos líneas celulares de cáncer, es decir, Paca2 y HepG2, usando REFS™ (véase la figura 22). Los números representan los cambios en veces en el nivel de expresión de proteínas particulares.

Para validar las conexiones proteicas identificadas usando el método anterior, los marcadores que se identificaron que estaban en las proximidades de LDHA en la red de simulación se introdujeron en IPA, un programa de software que usa redes neurales para determinar la unión molecular entre salidas experimentales a redes basándose en bibliografía publicada previamente. La salida del programa IPA se muestra en la figura 23, en la que se identificó que los marcadores en formas grises están en las proximidades de LDHA en la red de simulación generada por la plataforma y los marcadores en formas no rellenas son conexiones identificadas por IPA basándose en conocimiento conocido en bibliografía publicada previamente.

Se observó que los marcadores identificados en la salida de la tecnología de plataforma de biología de interrogación (mostrada en la figura 21), es decir, DHX9, HNRNPC, CKAP4, HSPA9, PARP1, HADHA, PHB2, ATP5A1 y CANX, estaban conectados a marcadores de cáncer bien conocidos como TP53 y PARK7 dentro de la red generada por IPA (mostrada en la figura 23). El hecho de que los factores identificados mediante el uso de la plataforma de biología de interrogación compartan conectividad con factores conocidos publicados en la bibliografía científica valida la precisión de la red creada mediante el uso de la plataforma de biología de interrogación. Además, la asociación de red dentro de la subred de LDHA creada mediante el uso de los resultados de la plataforma de biología de interrogación demostró la presencia de influencia direccional de cada factor, en contraste con la red de IPA en la que la unión entre entidades moleculares no proporciona direccionalidad funcional entre los nodos que

interaccionan. Por tanto, al emplear un enfoque sin sesgo para la generación, la integración y la ingeniería inversa de datos para crear un modelo computacional seguido de simulación y análisis de red diferencial, la plataforma de descubrimiento de biología de interrogación permite la comprensión de mecanismos hasta ahora desconocidos en la fisiopatología del cáncer que están en congruencia con el conocimientos científicos bien establecidos de la fisiopatología de la enfermedad.

La figura 19 muestra el efecto del tratamiento con CoQ10 sobre los ganglios posteriores (los números de registro de proteína Pubmed se enumeran en la figura 19) basándose en los datos de expresión de proteína de iProfiling. El número de registro de proteína P00338 es LDHA. La validación en laboratorio de investigación de los datos de proteómica se realizó para la expresión de LDHA en células HepG2 (véase la figura 20). Tal como se muestra en la figura 20, los niveles de expresión de LDHA disminuyeron cuando se trató HepG2 con CoQ10 50 uM o CoQ10 100 uM durante 24 ó 48 horas.

Para los marcadores de cáncer bien conocidos TP53, Bcl-2, Bax y caspasa3, se realizó la validación en laboratorio de investigación de los efectos del tratamiento con CoQ10 sobre el nivel de expresión de estos marcadores en las células SKMEL 28 (véase la figura 24 y la figura 25).

REALIZACIÓN DETALLADA 2: Empleo de la tecnología de plataforma para construir una red de cáncer delta-delta

En esta realización detallada, se empleó la tecnología de plataforma descrita con detalle anteriormente para integrar los datos obtenidos de un modelo de cáncer *in vitro* personalizado e identifican, de ese modo, nuevas proteínas/rutas que controlan la patogenia del cáncer. Los mapas relacionales resultantes de este análisis han proporcionado dianas de tratamiento del cáncer, así como marcadores de diagnóstico/pronóstico asociados con el cáncer.

En resumen, se sometieron cuatro líneas de cáncer (PaCa2, HepG2, PC3 y MCF7) y dos líneas de células normales (THLE2 y HDFa) a diversas condiciones que simulan un entorno experimentado por las células cancerosas *in vivo*. Específicamente, se expusieron las células por separado a cada una de las condiciones hiperglucémicas, condiciones hipóxicas y tratamiento con ácido láctico. Por ejemplo, se creó una condición hiperglucémica mediante el cultivo de las células en medios que contenían glucosa 22 mM. Se indujo una condición hipóxica al colocar las células en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que se inundó con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Para el tratamiento con ácido láctico, cada línea celular se trató con ácido láctico 0 ó 12,5 mM. Además de exponer las células a cada una de las tres condiciones anteriores por separado, también se expusieron las células a combinaciones de dos o las tres condiciones (es decir, condiciones hiperglucémicas e hipóxicas; condición hiperglucémica y ácido láctico; condición hipóxica y ácido láctico; y, condiciones hiperglucémica e hipóxica y ácido láctico).

El modelo celular que comprende las células mencionadas anteriormente, en el que cada tipo de célula se expuso a cada condición descrita anteriormente, se interrogó adicionalmente exponiendo las células a una perturbación ambiental mediante tratamiento con coenzima Q10. Específicamente, se trataron las células con coenzima Q10 a 0, 50 µM o 100 µM.

Las muestras de células, así como muestras de medios que contenían el secretoma de las células, para cada línea celular expuesta a cada condición (o combinación de condiciones), con y sin tratamiento con coenzima Q10, se recogieron en diversos momentos después del tratamiento, incluyendo después de 24 horas y 48 horas de tratamiento.

Además, se llevaron a cabo experimentos de interferencia entre dos células cancerosas diferentes, células PaCa2 y células HepG2, en las que se cocultivaron células PaCa2 y HepG2. Este enfoque de cocultivo se denomina experimento de secretoma extracelular (SEC). El primer sistema celular (PaCa2) se sembró en los insertos de los pocillos de una cámara de crecimiento de tipo Transwell. Generalmente se usaron placas de seis pocillos para permitir un mejor análisis estadístico. En el momento de la siembra del primer sistema celular en los insertos, se colocaron los insertos en una placa de 6 pocillos independiente. El segundo sistema celular (HepG2) se sembró en la bandeja primaria. La placa de 6 pocillos que contenía los insertos, que contenía el primer sistema celular, y la bandeja primaria que contenía el segundo sistema celular se incubaron a 37°C durante la noche. Cada uno de los sistemas celulares se hizo crecer en sus medios específicos de células respectivos (en los que, alternativamente, cada uno de los sistemas celulares podía hacerse crecer en un medio adaptado para soportar el crecimiento de ambos tipos de células). El segundo día, se administró el tratamiento predeterminado mediante intercambio de medios. Específicamente, los insertos que contenían el primer sistema celular y los medios respectivos del primer sistema celular se colocaron en la bandeja primaria que contenía el segundo sistema celular y los medios respectivos del segundo sistema celular. Sin embargo, en todos los casos de cocultivo, las células cocultivadas se habían expuesto a la misma "condición de cáncer" (por ejemplo, hiperglucemia, hipoxia, ácido láctico o combinaciones de las mismas), aunque por separado, durante el primer día anterior al cocultivo. Es decir, el primer sistema celular en los insertos y el segundo sistema celular en las bandejas se expusieron a la misma condición antes de trasladarse a una disposición de "cocultivo". Luego se incubó la bandeja durante un periodo de tiempo

predeterminado, por ejemplo, 24 horas o 48 horas. Se configuraron pocillos duplicados con las mismas condiciones, y se agruparon las células para producir material suficiente para el posterior análisis proteómico. Los medios que contenían el secretoma (alícuota de 1 ml), las células de los insertos y las células de los pocillos de la bandeja primaria se recogieron como muestras independientes. Los experimentos se realizaron por triplicado para proporcionar un mejor poder estadístico.

También se llevaron a cabo experimentos de interferencia mediante experimentos de “intercambio de medios”. Específicamente, se recogió un medio de cultivo o “secretoma” del primer sistema celular (PaCa2) después de 24 horas o 48 horas tras la perturbación y/o acondicionamiento y luego se añadió al segundo sistema celular durante 24-48 horas. Luego se recogió el medio de cultivo final o “secretoma” del segundo sistema celular. Todos los secretomas finales se sometieron a análisis proteómico.

Después de la exposición del sistema celular a las “condiciones de cáncer” descritas anteriormente, la perturbación (es decir, el tratamiento con coenzima Q10) y/o las condiciones producidas en el secretoma de una célula emparejada a partir de un experimento de cocultivo, se analizó entonces la respuesta de las células mediante el análisis de diversas lecturas del sistema celular. Las lecturas incluyeron datos proteómicos, específicamente expresión de proteínas intracelulares, así como proteínas secretadas en medios de cultivo celular, y datos funcionales, específicamente bioenergética celular.

Se realizó iProfiling de los cambios en la expresión de la proteína celular total mediante proteómica cuantitativa para muestras de células y medios recogidos para cada línea celular (líneas celulares normales y cancerosas) expuesta a cada condición (o combinación de condiciones), con o sin la “perturbación ambiental”, es decir, el tratamiento con coenzima Q10, usando las técnicas descritas anteriormente en la descripción detallada.

Además, se generaron perfiles bioenergéticos de cada línea celular (líneas celulares normales y de cáncer) expuesta a cada condición (o combinación de condiciones), con o sin la “perturbación ambiental”, es decir, el tratamiento con coenzima Q10, empleando el analizador Seahorse esencialmente según lo recomendado por el fabricante. Los electrodos registraron la tasa de consumo de oxígeno (TCO) y la tasa de acidificación extracelular (TAEC) en una cámara de 7 μ l creada con el cartucho empujando contra la placa de cultivo de Seahorse.

Los datos de proteómica recopilados para cada línea celular en cada condición/condiciones y con/sin cada perturbación, y los datos de perfiles bioenergéticos recopilados para cada línea celular en cada condición/condiciones y con/sin cada perturbación, se procesaron entonces mediante el sistema REFS™. Se generó una “red perturbada por el cáncer compuesta” a partir de datos combinados obtenidos a partir de todas las líneas celulares de cáncer, habiéndose expuesto cada una de ellas a cada condición específica (y combinación de condiciones), y se expuso adicionalmente a la perturbación (CoQ10). Se generó una “red no perturbada por el cáncer compuesta” a partir de datos combinados obtenidos de todas las líneas celulares de cáncer, habiéndose expuesto cada una de ellas a cada condición específica (y combinación de condiciones), sin perturbación (sin CoQ10). Del mismo modo, se generó una “red perturbada normal compuesta” a partir de datos combinados obtenidos de todas las líneas celulares normales, habiéndose expuesto cada una de ellas a cada condición específica (y combinación de condiciones), y se expuso adicionalmente a la perturbación (CoQ10). Se generó una “red no perturbada normal compuesta” a partir de datos combinados obtenidos de todas las líneas celulares normales, habiéndose expuesto cada una de ellas a cada condición específica (y combinación de condiciones), sin perturbación (sin CoQ10).

A continuación, se generaron “redes compuestas de simulación” (también denominadas en el presente documento “redes de simulación”) para cada una de las cuatro redes compuestas descritas anteriormente usando REFS™. Para lograr esto, se simuló cada nodo en la red compuesta de consenso dada (aumentando o disminuyendo 10 veces) para generar redes de simulación usando REFS™, tal como se describió con detalle anteriormente en la descripción detallada.

El área bajo la curva y los cambios en veces para cada arista que conecta un nodo primario a un nodo secundario en las redes de simulación los extrajo un programa personalizado que usa el lenguaje de programación R, en el que el lenguaje de programación R es un entorno de software de código abierto para grafos y computación estadística.

Finalmente, se generaron redes delta, en el que las redes delta representan el diferencial entre dos redes compuestas de simulación. Se generaron redes delta a partir de las redes compuestas de simulación. Para generar una red diferencial de cáncer frente a normal en respuesta a coenzima Q10 (red delta-delta), se realizaron etapas de comparación consecutivas tal como se ilustra en la figura 26, mediante un programa personalizado que usa el lenguaje de programación PERL.

En primer lugar, se compararon las redes no tratadas de cáncer (T0) y tratadas de cáncer (T1) usando el programa R, y se separaron las redes T1 tratadas de cáncer únicas (véase la forma de media luna de color gris oscuro en la figura 26). Esto representa la red “delta” de T1 de cáncer \cap (intersección) de T0 cáncer. La interacción/asociaciones proteicas dentro de esta red delta puede considerarse que representa la respuesta del cáncer única al tratamiento con coenzima Q10.

De manera similar se compararon, las redes no tratada normal (T0) y tratada normal (T1) usando el programa R, y se separaron las redes T1 tratadas normales únicas (véase la forma de media luna de color gris claro en la figura 26). Esto representa la red “delta” T1 normal \cap T0 normal. Puede considerarse que las interacciones/asociaciones proteicas dentro de esta red delta representan la respuesta de célula normal única al tratamiento con coenzima Q10.

Finalmente, se compararon redes de T1 de cáncer únicas (véase la forma de media luna de color gris oscuro en la figura 26) y redes de T1 normal únicas (véase la forma de media luna de color gris claro en la figura 26) usando el programa R, y se generaron redes que son únicas sólo para el cáncer, y no presentes en las células normales, en respuesta a la coenzima Q10 (véase la figura 26). Esta colección de interacciones/asociaciones proteicas representa las rutas únicas dentro de las células cancerosas que no están presentes en las células normales tras el tratamiento con coenzima Q10. Esta colección de interacciones/asociaciones proteicas se denomina “red delta-delta”, ya que es un mapa diferencial producido a partir de una comparación de un mapa diferencial de células cancerosas y un mapa diferencial de células de control normales.

La salida de los programas PERL y R se introdujo en Cytoscape, un programa de código abierto, para generar una representación visual de la red delta-delta.

Las redes delta-delta identificadas usando el método descrito en el presente documento son muy útiles para identificar dianas para el tratamiento del cáncer. Por ejemplo, según la red delta-delta presentada en la figura 27, la proteína A inhibe la TCO3 (una medida para la fosforilación oxidativa) y mejora la TAEC3 (una medida para la glicólisis). Dado que esta interacción es única en las células cancerosas (debido a que la red delta-delta ha restado cualquier interacción que está presente comúnmente en las células normales tras el tratamiento con coenzima Q10), se espera que la inhibición de la expresión de la proteína A reduzca el metabolismo energético basándose en la glicólisis, que es un sello distintivo de la ruta metabólica del cáncer, y desplazar las células hacia un metabolismo energético basado en fosforilación oxidativa, que es un fenotipo más asociado estrechamente con las células normales. Por tanto, se espera que una terapia combinada que use la coenzima Q10 y el inhibidor de la proteína A sea eficaz para tratar el cáncer, al menos en parte al desplazar el perfil de metabolismo energético de la célula cancerosa a lo que se asemeja a una célula normal.

La ventaja de la tecnología de plataforma de biología de interrogación de la invención se ilustra adicionalmente mediante el uso de un ejemplo sustantivo en el que se comparó una subred derivada de redes causales con una red molecular usando IPA, un programa de software que usa redes neurales para determinar la unión molecular entre salidas experimentales con redes basadas en bibliografía previamente publicada. La subred causal que contiene PARK7 generada usando la plataforma de biología de interrogación (mostrada en la figura 29) se usa como ejemplo sustantivo. Todas las firmas moleculares de la red de PARK7 de la plataforma de biología de interrogación se incorporaron a IPA para generar una red basada en evidencias bibliográficas conocidas/existentes. Luego se compararon las salidas de red entre la salida de biología de interrogación y la generada mediante el uso de IPA.

Se observó que seis marcadores identificados por la salida de la tecnología de plataforma de biología de interrogación (mostrada en la figura 29), es decir, A, B, C, X, Y y Z en las figuras 27-29, estaban conectados a TP53 dentro de la red generada por IPA (figura 28). Entre los seis marcadores, en la bibliografía se ha notificado que A, B y C están asociados con el cáncer, así como con HSPA1A/HSPA1B. X, Y y Z se identificaron como “centros” o controladores clave del estado de cáncer y, por tanto, se identifican como nuevos marcadores del cáncer. Además, MIF1 y KARS también se identificaron como “centros” o controladores clave del estado de cáncer y, por tanto, se identifican como nuevos marcadores del cáncer. El hecho de que los factores identificados mediante el uso de la plataforma de biología de interrogación compartan conectividad con factores conocidos publicados en la bibliografía científica valida la precisión de la red creada mediante el uso de la plataforma de biología de interrogación. Además, la asociación de red dentro de la subred de PARK7 creada mediante el uso de las salidas de la plataforma de biología de interrogación (mostrada en la figura 29) demostró la presencia de influencia direccional de cada factor, en contraste con la red de IPA (mostrada en la figura 28) en el que la unión entre entidades moleculares no proporciona direccionalidad funcional entre los nodos que interaccionan. Además, las salidas de la plataforma de biología de interrogación (mostrada como líneas de puntos en la figura 29) demostraron la asociación de estos componentes que conducen a un mecanismo potencial a través de PARK7. Se observó que la proteína C, la proteína A y otros nodos de PARK7 son controladores clave del metabolismo del cáncer (figura 27).

Tal como demuestra la presente realización detallada, al emplear un enfoque sin sesgo para la generación, la integración y la ingeniería inversa de datos para crear un modelo computacional seguido de simulación y análisis de red diferencial, la plataforma de descubrimiento de biología de interrogación permite la comprensión de mecanismos hasta ahora desconocidos en la fisiopatología del cáncer que están en congruencia con una comprensión científica bien establecida de la fisiopatología de la enfermedad.

REALIZACIÓN DETALLADA 3: Empleo de la tecnología de plataforma para construir una red delta-delta de diabetes/obesidad/enfermedad cardiovascular

En esta realización detallada, se empleó la tecnología de plataforma descrita con detalle anteriormente en la

descripción detallada para integrar datos obtenidos de un modelo personalizado de diabetes/obesidad/enfermedad cardiovascular (ECV), y para identificar nuevas proteínas/rutas que controlan la patogenia de la diabetes/obesidad/ECV. Los mapas relacionales resultantes de este análisis han proporcionado dianas de tratamiento de diabetes/obesidad/ECV, así como marcadores de diagnóstico/pronóstico asociados con diabetes/obesidad/ECV.

Se sometieron cinco líneas celulares humanas primarias, a saber, adipocitos, miotubos, hepatocitos, células de músculo liso aórtico (HASMC) y células tubulares proximales (HK2) a una de las cinco condiciones que simulan un entorno experimentado por estas células relevantes para la enfermedad *in vivo*. Específicamente, cada una de las cinco líneas celulares se expuso por separado a cada una de las siguientes condiciones: condiciones hiperglucémicas, condiciones hiperlipidémicas, condiciones hiperinsulinémicas, condiciones hipóxicas y exposición a ácido láctico. La condición hiperglucémica se indujo mediante el cultivo de células en medios que contenían glucosa 22 mM. La condición hiperlipidémica se indujo mediante el cultivo de las células en medios que contenían palmitato de sodio 0,15 mM. La condición hiperinsulinémica se indujo mediante el cultivo de las células en medios que contenían insulina 1000 nM. La condición hipóxica se indujo al colocar las células en una cámara de incubador modular (MIC-101, Billups-Rothenberg Inc. Del Mar, CA), que se inundó con una mezcla de gases industriales que contenía el 5% de CO₂, el 2% de O₂ y el 93% de nitrógeno. Cada línea celular también se trató con ácido láctico 0 o 12,5 mM.

Además, se llevaron a cabo experimentos de interferencia entre dos pares diferentes de células, HASMC (sistema celular 1) y células HK2 (sistema celular 2) o células hepáticas (sistema celular 1) y adipocitos (sistema celular 2) en los que se cocultivaron las células emparejadas. Este enfoque de cocultivo se denomina experimento de secretoma extracelular (SEC). El primer sistema celular (por ejemplo, HASMC) se sembró en primer lugar en los insertos de los pocillos de una cámara de crecimiento de tipo Transwell. Se usaron placas de seis pocillos para permitir un mejor análisis estadístico. En el momento de la siembra con el primer sistema celular en los insertos, se colocaron los insertos en una placa de 6 pocillos independiente. El segundo sistema celular (por ejemplo, HK2) se sembró en la bandeja primaria. La bandeja de insertos que contenía el primer sistema celular y la bandeja primaria que contenía el segundo sistema celular se incubaron a 37°C durante la noche. Cada uno de los sistemas celulares se hizo crecer en los medios específicos de células específicas (en los que, alternativamente, cada uno de los sistemas celulares podía hacerse crecer en un medio adaptado para soportar el crecimiento de ambos tipos de células). El segundo día, se administró el tratamiento predeterminado mediante intercambio de medios. Específicamente, se colocaron los insertos que contenían el primer sistema celular en la bandeja primaria que contenía el segundo sistema celular. Luego se incubó la bandeja durante un periodo de tiempo predeterminado, por ejemplo, 24 horas o 48 horas. Se configuraron pocillos duplicados con las mismas condiciones, y se agruparon las células para producir material suficiente para el análisis 2D. Los medios (alícuota de 1 ml), las células de los insertos y las células de los pocillos de la bandeja primaria se recogieron como muestras independientes. Los experimentos se realizaron por triplicado para proporcionar un mejor poder de análisis estadístico.

También se llevaron a cabo experimentos de interferencia mediante experimentos de "intercambio de medios". Específicamente, un medio de cultivo o "secretoma" del primer sistema celular, HASMC se recogió después de 24 horas o 48 horas después de la perturbación o acondicionamiento y luego se añadió al segundo sistema celular, adipocitos, durante 24-48 horas. Luego se recogió el medio de cultivo final o "secretoma" del segundo sistema celular. Todos los secretomas finales se sometieron a análisis proteómico.

El modelo celular que comprendía las células mencionadas anteriormente, en el que se expusieron las células a cada condición descrita anteriormente, se "interrogó" adicionalmente exponiendo las células a una "perturbación ambiental" mediante tratamiento con coenzima Q10. Específicamente, se trataron las células con coenzima Q10 a 0, 50 µM o 100 µM.

Se recogieron muestras de células para cada línea celular, condición y tratamiento con coenzima Q10 en diversos momentos después del tratamiento, incluyendo después de 24 horas y 48 horas de tratamiento. Para determinadas células y en determinadas condiciones, también se recogieron y analizaron muestras de medios.

Se realizó iProfiling de los cambios en la expresión de la proteína celular total mediante proteómica cuantitativa para muestras de células y medios recogidas para cada línea celular en cada condición y con cada "perturbación ambiental", es decir, tratamiento con coenzima Q10, usando las técnicas descritas anteriormente en la descripción detallada.

Los datos de proteómica recopilados para cada línea celular enumerados anteriormente en cada condición y con cada perturbación, y los datos de perfiles bioenergéticos recopilados para cada línea celular en cada condición y con cada perturbación, se procesaron mediante el sistema REFS™. Se generó una red perturbada compuesta a partir de datos combinados obtenidos de todas las líneas celulares para una condición específica (por ejemplo, hiperglucemia) expuesta a perturbación (CoQ10). Se generó una red no perturbada compuesta a partir de datos combinados obtenidos de todas las líneas celulares para la misma condición específica (por ejemplo, hiperglucemia), sin perturbación (sin CoQ10). De manera similar, se generó una red perturbada compuesta a partir de datos combinados obtenidos de todas las líneas celulares durante una segunda condición de control (por

ejemplo, glucemia normal) expuesta a perturbación (CoQ10). Se generó una red no perturbada compuesta a partir de datos combinados obtenidos de todas las líneas celulares para el mismo segundo estado de control (por ejemplo, glucemia normal), sin perturbación (sin CoQ10).

5 Cada nodo en las redes compuestas de consenso descritas anteriormente se simuló (aumentando o disminuyendo 10 veces) para generar redes de simulación usando REFS™, tal como se describió con detalle anteriormente en la descripción detallada.

10 El área bajo la curva y los cambios en veces para cada arista que conecta un nodo primario a un nodo secundario en las redes de simulación los extrajo un programa personalizado que usa el lenguaje de programación R, en el que el lenguaje de programación R es un entorno de software de código abierto para grafos y computación estadística.

15 Se generaron redes delta a partir de las redes compuestas simuladas. Para generar una red diferencial de condición de diabetes/obesidad/enfermedad cardiovascular frente a condición normal en respuesta a la coenzima Q10 (red delta-delta), se realizaron etapas de comparación tal como se ilustra en la figura 30, mediante un programa personalizado que usa el lenguaje de programación PERL.

20 Específicamente, tal como se muestra en la figura 30, el tratamiento T1 se refiere al tratamiento con coenzima Q10 y NG y HG se refieren a condiciones normales y de hiperglucemia. Las aristas únicas de NG en la red delta NG∩HG se compararon con las aristas únicas de HGT1 en la red delta HG∩HGT1. Las aristas en la intersección de NG y HGT1 son aristas HG que se restauran a NG con T1. Las aristas HG restauradas a NG con T1 se superpusieron a la red delta NG∩HG (mostrada en círculos de colores más oscuros en la figura 31)

25 Específicamente, se comparó un mapa compuesto simulado de la condición de glucemia normal (NG) y un mapa compuesto simulado de la condición de hiperglucemia (HG) usando un programa Perl personalizado para generar aristas únicas de la condición de glucemia normal. Se compararon un mapa compuesto simulado de la condición de hiperglucemia sin tratamiento con coenzima Q10 (HG) y un mapa simulado de condición de hiperglucemia con tratamiento con coenzima Q10 (HGT1) usando un programa Perl personalizado para generar aristas únicas de la condición de hiperglucemia con tratamiento con coenzima Q10 (HGT1). Las aristas en la intersección de las aristas
30 únicas de la condición de glucemia normal (NG) y las aristas únicas de la condición de hiperglucemia con el tratamiento con coenzima Q10 (HGT1) se identificaron usando el programa Perl. Estas aristas representan factores/redes que se restauran a la condición de glucemia normal a partir de la condición de hiperglucemia mediante el tratamiento de la coenzima Q10. La red delta-delta de aristas de hiperglucemia restauradas a la normalidad con el tratamiento con coenzima Q10 se superpuso a la red delta de glucemia normal ∩ hiperglucemia.
35 En la figura 31 se muestra una muestra de las redes superpuestas. La figura 31 es una red diferencial de condición de diabetes/obesidad/enfermedad cardiovascular frente a condición normal a modo de ejemplo en respuesta a la coenzima Q10 (red delta-delta). Los círculos de color más oscuro en la figura 31 son aristas identificadas que se restauraron a una condición de glucemia normal a partir de una condición de hiperglucemia mediante el tratamiento de la coenzima Q10. Los círculos de colores más claros en la figura 31 se identifican con aristas de hipercemia normal únicas.

La salida de los programas PERL y R se introdujeron en Cytoscape, un programa de código abierto, para generar una representación visual de la red delta-delta.

45 De manera similar a los experimentos descritos anteriormente para la hiperglucemia frente a la condición glucémica normal, se compararon una red compuesta simulada de condición de hiperlipidemia (que combina datos de todas las células relacionadas con diabetes/obesidad/cardiovascular descritas anteriormente) sin el tratamiento con coenzima Q10 y una red compuesta simulada de condición de hiperlipidemia (combinando datos de todas las células relacionadas con diabetes/obesidad/cardiovascular, descritas anteriormente) con el tratamiento con coenzima Q10
50 usando el programa Perl para generar aristas únicas de la condición de hiperlipidemia con el tratamiento con coenzima Q10. Las aristas en la intersección de las aristas únicas de la condición de lipidemia normal y las aristas únicas de la condición hiperlipidémica con el tratamiento con coenzima Q10 se identificaron usando el programa Perl. Estas aristas representan factores/redes que se restauran a una condición de lipidemia normal a partir de una condición de hiperlipidemia mediante el tratamiento de la coenzima Q10. Una red delta-delta de aristas de hiperlipidemia restauradas a la normalidad con el tratamiento con coenzima Q10 se superpuso a la red delta de lipidemia normal ∩ hiperlipidemia. Una muestra de las redes superpuestas se muestra en la figura 32. Los círculos de color más oscuro en la figura 32 son aristas identificadas que se restauraron a una condición de lipidemia normal a partir de una condición de hiperlipidemia mediante el tratamiento de la coenzima Q10. Los círculos de colores más
55 claros en la figura 32 se identifican como aristas de lipidemia normal únicas. Se identificó FASN como un factor importante de una ruta de señalización que modula el efecto de la coenzima Q10 de restaurar la hiperlipidemia a una condición de lipidemia normal.
60

65 Las enzimas de síntesis de ácidos grasos, las ácido graso sintasas como FASN se han implicado en casi todos los aspectos de las alteraciones metabólicas humanas, tales como obesidad, resistencia a la insulina o dislipidemia. Los inhibidores de FASN se han propuesto como moléculas iniciales para el tratamiento de la obesidad, aunque se desconocen los mecanismos moleculares (Mobbs *et al* 2002). Se ha demostrado que la cerulenina y los inhibidores

del compuesto sintético C75-FASN tienen un efecto en la reducción de la ingesta de alimentos y en la pérdida de peso (Loftus *et al* 2000).

5 El hecho de que FASN se identificase mediante la tecnología de plataforma descrita en el presente documento como un factor importante en la ruta de señalización que modula el efecto de la coenzima Q10 de restaurar un estado diabético a uno normal, tal como se muestra en la figura 32, validó la precisión de esta red delta-delta. Por tanto, otros factores nuevos identificados en esta red delta-delta serán posibles factores terapéuticos u dianas farmacológicas para investigación adicional.

10 Realización detallada 4: Empleo de la tecnología de plataforma para construir modelos de cardiotoxicidad inducida por fármacos

15 En esta realización detallada, se empleó la tecnología de plataforma descrita con detalle anteriormente en la descripción detallada para integrar los datos obtenidos de un modelo de cardiotoxicidad personalizado, y para identificar nuevas proteínas/rutas que controlan la patogenia/toxicidad de los fármacos. Los mapas relacionales resultantes de este análisis han proporcionado biomarcadores de toxicidad.

20 En el corazón sano, la función contráctil depende de un equilibrio de oxidación de hidratos de carbono y ácidos grasos. Se cree que el desequilibrio crónico en la captación, utilización, biogénesis y secreción por orgánulos en tejido no adiposo (corazón e hígado) está en el centro del daño y la disfunción mitocondrial y es un actor clave en la cardiotoxicidad inducida por fármacos. En el presente documento, los solicitantes describen un enfoque de sistemas que combina las firmas de proteína y lípidos con ensayos de punto final funcionales que analizan específicamente la bioenergética celular y la función de la membrana mitocondrial. Los modelos *in vitro* que comprenden cardiomiocitos de diabéticos y normales complementados con ácido graso en exceso e hiperglucemia se trataron con un panel de fármacos para crear firmas y posibles mecanismos de toxicidad. Los solicitantes demostraron los variados efectos de los fármacos en la desestabilización de las mitocondrias al perturbar el componente del metabolismo energético en diversos niveles, incluyendo (i) la desregulación de las redes transcripcionales que controlan la expresión de los genes del metabolismo energético mitocondrial; (ii) la inducción de GPAT1 y tafazzina en cardiomiocitos de diabéticos, iniciándose de ese modo la síntesis y remodelación de fosfolípidos *de novo* en la membrana mitocondrial; 30 y (iii) el destino alterado de ácidos grasos en los cardiomiocitos de diabéticos, que influye en la captación, oxidación de ácidos grasos y síntesis de ATP. Además, los solicitantes combinaron el poder de la biología de laboratorio de investigación y la plataforma de minería de datos basada en IA para generar una red causal basándose en modelos bayesianos. Se usaron redes de proteínas y lípidos que son causales para la pérdida de la función celular normal para discernir los mecanismos de toxicidad inducida por fármacos de los mecanismos de protección celular. Este nuevo enfoque servirá como nueva herramienta poderosa para comprender el mecanismo de toxicidad mientras 35 permite el desarrollo de productos terapéuticos más seguros que corrigen un fenotipo alterado.

40 Se sometieron cardiomiocitos humanos a condiciones que simulaban un entorno de diabéticos experimentado por las células relevantes para la enfermedad *in vivo*. Específicamente, se expusieron las células a condiciones hiperglucémicas y condiciones de hiperlipidemia. La condición hiperglucémica se indujo mediante el cultivo de células en medios que contenían glucosa 22 mM. La condición de hiperlipidemia se indujo mediante el cultivo de las células en medios que contenían L-carnitina 1 mM, ácido oleico 0,7 mM y ácido linoleico 0,7 mM.

45 El modelo celular que comprendía las células mencionadas anteriormente, en el que se expusieron las células a cada condición descrita anteriormente, se "interrogó" adicionalmente exponiendo las células a una "perturbación ambiental" mediante tratamiento con un fármaco para la diabetes (T) que se sabe que provoca cardiotoxicidad, una molécula de rescate (R) o tanto el fármaco para la diabetes como molécula de rescate (T + R). Específicamente, se trataron las células con fármaco para la diabetes; o se trataron con la molécula de rescate, coenzima Q10 a 0, 50 μ M o 100 μ M; o se trataron tanto con el fármaco para la diabetes como con la molécula de rescate, coenzima Q10. 50

Se recogieron muestras de células de cada condición con cada tratamiento de perturbación en diversos momentos después del tratamiento, incluyendo después de 6 horas de tratamiento. Para determinadas condiciones, también se recogieron y analizaron muestras de medios.

55 Se realizó iProfiling de los cambios en la expresión de la proteína celular total mediante proteómica cuantitativa para las muestras de células y medios recogidas para cada condición y con cada "perturbación ambiental", es decir, tratamiento farmacológico para la diabetes, tratamiento con coenzima Q10 o ambos, usando las técnicas descritas anteriormente en la descripción detallada. Se llevaron a cabo experimentos de obtención de perfiles transcripcionales usando el sistema de amplificación cfx-384 de Biorad. Después de la recopilación de datos (Ct), se determinó el cambio en veces final con respecto al control usando el método δ Ct tal como se describe en el protocolo del fabricante. Se llevaron a cabo experimentos de lipidómica usando espectrometría de masas. Los ensayos funcionales tales como la tasa de consumo de oxígeno TCO se midieron empleando el analizador Seahorse esencialmente según lo recomendado por el fabricante. Los electrodos registraron la TCO en una cámara de 7 μ l creada con el cartucho empujando contra la placa de cultivo de Seahorse. 60

65 Tal como se muestra en la figura 35, la red transcripcional y la expresión de genes del metabolismo energético

mitocondrial humano en cardiomiocitos de diabéticos (cardiomiocitos acondicionados en hiperglucemia e hiperlipidemia) se compararon entre tratamientos perturbados y no perturbados. Específicamente, se compararon datos de la red transcripcional y la expresión de genes del metabolismo energético mitocondrial humano entre cardiomiocitos de diabéticos tratados con el fármaco para la diabetes (T) y las muestras de cardiomiocitos de diabéticos no tratados (UT). Los datos de la red transcripcional y la expresión de genes del metabolismo energético mitocondrial humano se compararon entre los cardiomiocitos de diabéticos tratados con el fármaco diabético y la molécula de rescate, coenzima Q10 (T + R) y muestras de cardiomiocitos de diabéticos no tratados (UT). En comparación con los datos de los cardiomiocitos de diabéticos no tratados, se alteraron la expresión y transcripción de determinados genes cuando los cardiomiocitos de diabéticos se trataron con un fármaco para la diabetes. Se demostró que la molécula de rescate, coenzima Q10 revierte el efecto tóxico del fármaco para la diabetes y normaliza la expresión y transcripción génica.

Tal como se muestra en la figura 36A, se cultivaron cardiomiocitos en condiciones de normoglucemia (NG) o hiperglucemia (HG) y se trataron con un fármaco para la diabetes solo (T) o tanto con el fármaco para la diabetes como con la molécula de rescate, coenzima Q10 (T + R). Los niveles de expresión de proteína de GPAT1 y TAZ para cada condición y cada tratamiento se sometieron a prueba con inmunotransferencia de tipo Western. Tanto GPAT1 como TAZ estaban regulados por incremento en cardiomiocitos acondicionados en hiperglucemia y tratados con fármaco para la diabetes. Cuando se trataron cardiomiocitos acondicionados en hiperglucemia tanto con el fármaco para la diabetes como con la molécula de rescate, coenzima Q10, se normalizó el nivel de expresión de proteína regulado por incremento de GPAT1 y TAZ.

Tal como se muestra en la figura 37A, se llevaron a cabo experimentos de la tasa de consumo de oxígeno mitocondrial (%) para muestras de cardiomiocitos acondicionados en hiperglucemia. Los cardiomiocitos acondicionados en hiperglucemia o bien no se trataron (UT), se trataron con el fármaco para la diabetes T1 que se sabe que provoca cardiotoxicidad, se trataron con el fármaco para la diabetes T2 que se sabe que provocan cardiotoxicidad, se trataron tanto con el fármaco para la diabetes T1 como con la molécula de rescate, coenzima Q10 (T1 + R), o bien se trataron tanto con el fármaco para la diabetes T2 como la molécula de rescate, coenzima Q10 (T2 + R). En comparación con las muestras de control no tratadas, la TCO mitocondrial disminuyó cuando los cardiomiocitos acondicionados en hiperglucemia se trataron con el fármaco para la diabetes T1 o T2. Sin embargo, la TCO mitocondrial se normalizó cuando los cardiomiocitos acondicionados en hiperglucemia se trataron tanto con el fármaco para la diabetes como con la molécula de rescate, coenzima Q10 (T1 + R o T2 + R).

Tal como se muestra en la figura 37B, se llevaron a cabo experimentos de síntesis de ATP mitocondrias para muestras de cardiomiocitos acondicionados en hiperglucemia. Los cardiomiocitos acondicionados en hiperglucemia o bien no se trataron (UT), se trataron con un fármaco para la diabetes (T) o bien se trataron tanto con el fármaco para la diabetes como con la molécula de rescate, coenzima Q10 (T + R). En comparación con las muestras de control no tratadas, la síntesis de ATP mitocondrial se reprimió cuando los cardiomiocitos acondicionados en hiperglucemia se trataron con fármaco para la diabetes (T).

Tal como se muestra en la figura 38, basados en los datos proteómicos recopilados, las proteínas reguladas por disminución mediante el tratamiento farmacológico se anotaron con términos GO. Las proteínas implicadas en el metabolismo energético mitocondrial estaban reguladas por disminución cuando los cardiomiocitos acondicionados en hiperglucemia se trataron con un fármaco para la diabetes que se sabe que provoca cardiotoxicidad.

El sistema REFS™ procesó luego los datos de proteómica, lipidómica, perfiles transcripcionales, ensayos funcionales y de inmunotransferencia de tipo Western recopilados para cada condición y con cada perturbación. Se generaron redes perturbadas compuestas a partir de datos combinados obtenidos de una condición específica (por ejemplo, hiperglucemia o hiperlipidemia) expuesta a cada perturbación (por ejemplo, fármaco para la diabetes, CoQ10 o ambos). Se generaron redes no perturbadas compuestas a partir de datos combinados obtenidos de la misma condición específica (por ejemplo, hiperglucemia o hiperlipidemia), sin perturbaciones (no tratadas). De manera similar, se generaron redes perturbadas compuestas a partir de datos combinados obtenidos para una segunda condición de control (por ejemplo, glucemia normal) expuesta a cada perturbación (por ejemplo, fármaco para la diabetes, CoQ10 o ambos). Se generaron redes no perturbadas compuestas a partir de datos combinados obtenidos de la misma segunda condición de control (por ejemplo, glucemia normal), sin perturbación (no tratadas).

Cada nodo en las redes compuestas de consenso descritas anteriormente se simuló (aumentando o disminuyendo 10 veces) para generar redes de simulación usando REFS™, tal como se describió con detalle anteriormente en la descripción detallada.

El área bajo la curva y los cambios en veces para cada arista que conecta un nodo primario a un nodo secundario en las redes de simulación los extrajo un programa personalizado que usa el lenguaje de programación R, en el que el lenguaje de programación R es un entorno de software de código abierto para grafos y computación estadística.

Se generaron redes delta a partir de las redes compuestas simuladas. Para generar una red diferencial de condición de toxicidad inducida por fármacos frente a condición normal en respuesta al fármaco para la diabetes (red delta), se realizaron etapas de comparación tal como se ilustra en la figura 39, mediante un programa personalizado que usa

el lenguaje de programación PERL.

5 Específicamente, tal como se muestra en la figura 39, UT se refiere a redes de expresión de proteína de cardiomiocitos de control no tratados en condición de hiperglucemia. El tratamiento T se refiere a las redes de expresión de proteína de cardiomiocitos tratados con fármaco para la diabetes en condiciones de hiperglucemia. Las aristas únicas de T en la red delta UT∩T se presentan en la figura 40.

10 Específicamente, se comparó un mapa compuesto simulado de cardiomiocitos no tratados en condición de hiperglucemia y un mapa compuesto simulado de cardiomiocitos tratados con fármaco para la diabetes en condición de hiperglucemia usando un programa Perl personalizado para generar aristas únicas de los cardiomiocitos tratados con fármaco para la diabetes en condición de hiperglucemia. La salida de los programas PERL y R se introdujo en Cytoscape, un programa de código abierto, para generar una representación visual de la red delta. Tal como se muestra en la figura 40, la red representa redes delta que están controladas por modelos no tratados frente a fármaco para la diabetes en cardiomiocitos/cardiotox. en condición de hiperglucemia.

15 A partir de la red diferencial de condición de toxicidad inducida por fármacos frente a condición normal mostrada en la figura 40, se identificaron proteínas que controlan la fisiopatología de la cardiotoxicidad inducida por fármacos, tales como GRP78, GRP75, TIMP1, PTX3, HSP76, PDIA4, PDIA1, CA2D1. Estas proteínas pueden funcionar como biomarcadores para la identificación de otros fármacos inductores de cardiotoxicidad. Estas proteínas también pueden funcionar como biomarcadores para la identificación de agentes que pueden aliviar la cardiotoxicidad.

20 Los experimentos descritos en esta realización detallada demuestran que la biología de membrana perturbada y el destino alterado de ácidos grasos libres en cardiomiocitos de diabéticos expuestos a tratamiento farmacológico representan la pieza central de la toxicidad inducida por fármacos. La integración de datos y la biología de red han permitido una mejor comprensión de la cardiotoxicidad y la identificación de nuevos biomarcadores predictivos de cardiotoxicidad.

Incorporación por referencia

30 La práctica de la presente invención empleará, a menos que se indique lo contrario, técnicas convencionales de formulación de proteínas, que se conocen bien en la técnica.

Equivalentes

35 Por tanto, las realizaciones anteriores deben considerarse en todos los aspectos ilustrativas más que limitativas de la invención descrita en el presente documento. Por tanto, el alcance de la invención está indicado por las reivindicaciones adjuntas más que por la descripción anterior, y todos los cambios que entran dentro del significado y el rango de equivalentes de las reivindicaciones, por tanto, se pretende que estén abarcados en el presente documento.

40

REIVINDICACIONES

1. Método para identificar un modulador de un proceso de enfermedad, comprendiendo dicho método:
 - 5 (1) establecer un modelo de enfermedad para el proceso de enfermedad, comprendiendo el modelo de enfermedad células de la enfermedad o células asociadas a enfermedad;
 - 10 (2) obtener un primer grupo de datos del modelo de enfermedad, en el que el primer grupo de datos representa niveles de expresión de una pluralidad de genes en las células de la enfermedad o células asociadas a enfermedad;
 - 15 (3) obtener un segundo grupo de datos del modelo de enfermedad, en el que el segundo grupo de datos representa una actividad funcional o una respuesta celular de las células de la enfermedad o las células asociadas a enfermedad;
 - 20 (4) generar una red de relaciones causales de consenso basándose en un conjunto de redes bayesianas de relaciones causales entre los niveles de expresión de la pluralidad de genes y la actividad funcional o respuesta celular basándose únicamente en el primer grupo de datos y el segundo grupo de datos usando un dispositivo de computación programado, en el que la generación de la red de relaciones causales de consenso no se basa en ninguna relación biológica conocida distinta del primer grupo de datos y el segundo grupo de datos;
 - 25 (5) generar una red de relaciones causales diferencial a partir de la comparación de la red de relaciones causales de consenso basándose en datos del modelo de enfermedad con una red de relaciones causales de comparación basándose en datos de célula de comparación de las células de comparación; y
 - 30 (6) identificar una relación causal única en el proceso de enfermedad como parte de la red de relaciones causales diferencial que está presente de manera única en la red de relaciones causales de consenso basándose en datos del modelo de enfermedad, incluyendo las células de la enfermedad o las células asociadas a enfermedad y ausente en la red de relaciones causales de comparación basándose en los datos de célula de comparación de las células de comparación, en el que la relación causal identificada única en el proceso de enfermedad está conectada a un nodo primario que representa una variable de entrada y un nodo secundario que representa una variable de entrada, en el que el nodo primario y el nodo secundario están conectados por una arista de un grafo acíclico dirigido que representa el modelo de enfermedad, en el que la variable de entrada para el nodo primario o la variable de entrada para el nodo secundario corresponde a un gen, que es un gen asociado con la relación causal única identificada, y en el que el gen asociado con la relación causal única se identifica como modulador del proceso de enfermedad,
 - 35 en el que el conjunto de redes bayesianas en las que se basa la red de relaciones causales de consenso incluye información direccional probabilística cuantitativa referente a las relaciones causales entre los niveles de expresión de la pluralidad de genes y la actividad funcional o respuesta celular;
 - 40 en el que el proceso de enfermedad es cáncer, diabetes, obesidad o enfermedad cardiovascular;
 - 45 en el que las células de la enfermedad o las células asociadas a enfermedad en el modelo de enfermedad están sujetas a uno o más estímulos ambientales relevantes para la enfermedad y están sujetas a una perturbación ambiental, que es cualquier estímulo físico y/o químico externo que pueda afectar a la función celular;
 - 50 en el que los datos de célula de comparación son de células de comparación que están sujetas a uno o más estímulos ambientales relevantes para la enfermedad y no están sujetas a la perturbación ambiental; y en el que las células de comparación son células no tratadas o células no perturbadas.
- 55 2. Método según la reivindicación 1, en el que el cáncer es cáncer de pulmón, cáncer de mama, cáncer de próstata, melanoma, carcinoma de células escamosas, cáncer colorrectal, cáncer de páncreas, cáncer de tiroides, cáncer de endometrio, cáncer de vejiga, cáncer de riñón, un tumor sólido, leucemia, linfoma no Hodgkin o un cáncer resistente a fármacos.
- 60 3. Método según la reivindicación 1, en el que el modulador estimula o promueve el proceso de enfermedad.
4. Método según la reivindicación 1, en el que el modulador inhibe el proceso de enfermedad.
- 65 5. Método según la reivindicación 4, en el que el modulador cambia la ruta metabólica de energía específicamente en las células de la enfermedad de una ruta glucolítica a una ruta de fosforilación oxidativa.

6. Método según la reivindicación 1, en el que el modelo de enfermedad comprende un cultivo *in vitro* de las células de la enfermedad o las células asociadas a enfermedad sujetas a una perturbación ambiental, que es cualquier estímulo físico y/o químico externo que pueda afectar a la función celular, y se obtienen los datos de célula de comparación a partir de un cultivo *in vitro* de células de la enfermedad idénticas o células asociadas a enfermedad no sujetas a la perturbación ambiental.
7. Método según la reivindicación 6, en el que la perturbación ambiental comprende uno o más de un contacto con un agente, un cambio en la condición de cultivo, una modificación/mutación genética introducida, y un vehículo (por ejemplo, vector) que provoca una modificación/mutación genética.
8. Método según la reivindicación 1, en el que el modelo de enfermedad representa un aspecto característico del proceso de enfermedad que comprende una condición de hipoxia, una condición hiperglucémica, una condición de cultivo rica en ácido láctico, una condición hiperlipidémica, una condición hiperinsulinémica, o combinaciones de las mismas.
9. Método según la reivindicación 1, en el que los niveles de expresión de una pluralidad de genes comprenden niveles de expresión de proteína y/o ARNm de la pluralidad de genes.
10. Método según la reivindicación 1, en el que el primer grupo de datos comprende además uno o más datos lipídicos, datos metabólicos, datos transcriptómicos y datos de polimorfismo de un solo nucleótido (SNP); y
- en el que la red de relaciones causales de consenso incluye además relaciones entre variables que representan el uno o más de datos lipídicos, datos metabólicos, datos transcriptómicos y datos de polimorfismo de un solo nucleótido (SNP), y las variables que representan niveles de expresión de la pluralidad de genes, y la actividad funcional o respuesta celular.
11. Método según la reivindicación 1, en el que el segundo grupo de datos comprende uno o más de perfiles bioenergéticos, proliferación celular, apoptosis, función de orgánulo y una asociación genotipo-fenotipo actualizada mediante modelos funcionales seleccionados de adenosina trifosfato (ATP), especies reactivas de oxígeno (ROS), fosforilación oxidativa (OXPHOS) y ensayos de Seahorse.
12. Método según la reivindicación 1, en el que la red de relaciones causales de consenso incluye además relaciones entre variables que representan el uno o más de datos de perfiles bioenergéticos, datos de proliferación celular, datos de apoptosis, datos de función de orgánulo y datos de asociación genotipo-fenotipo, y las variables que representan niveles de expresión de la pluralidad de genes, la actividad funcional o la respuesta celular.
13. Método según la reivindicación 1, en el que la etapa (4) se lleva a cabo mediante una plataforma de informática basada en inteligencia artificial (IA).
14. Método según la reivindicación 13, en el que la plataforma de informática basada en IA recibe toda la entrada de datos del primer grupo de datos y el segundo grupo de datos sin aplicar un punto de corte estadístico.
15. Método según la reivindicación 1, que comprende además, antes de la etapa 5, generar la relación causal de consenso de comparación basándose en los datos obtenidos de las células de comparación.
16. Método según la reivindicación 1, que comprende además validar la relación causal única identificada en un sistema biológico mediante la realización de uno o más experimentos basados en células o basados en animales.
17. Método según la reivindicación 1, en el que el método comprende además:
- generar una red de relaciones causales de consenso a partir de los primeros datos de célula de control, en el que los primeros datos de célula de control se obtienen de células que coinciden con las células de la enfermedad o las células asociadas a enfermedad que no están sujetas a uno o más estímulos ambientales relevantes para la enfermedad y que están sujetas a la perturbación ambiental;
- generar una red de relaciones causales de consenso a partir de los segundos datos de célula de control, en el que los segundos datos de célula de control se obtienen de células que coinciden con las células de la enfermedad o las células asociadas a enfermedad que no están sujetas a uno o más estímulos ambientales relevantes para la enfermedad y que no están sujetas a la perturbación ambiental; y
- generar una segunda red diferencial a partir de la comparación de la red de relaciones causales de

consenso basándose en los primeros datos de célula de control y la red de relaciones causales de consenso basándose en los segundos datos de célula de control;

- 5 en el que identificar una relación causal única en el proceso de enfermedad como parte de la red de relaciones causales diferencial incluye formar una red delta-delta a partir de la comparación de la red diferencial y la segunda red diferencial, y en el que la red delta-delta incluye relaciones que están presentes en la red diferencial y ausentes en la segunda red diferencial.

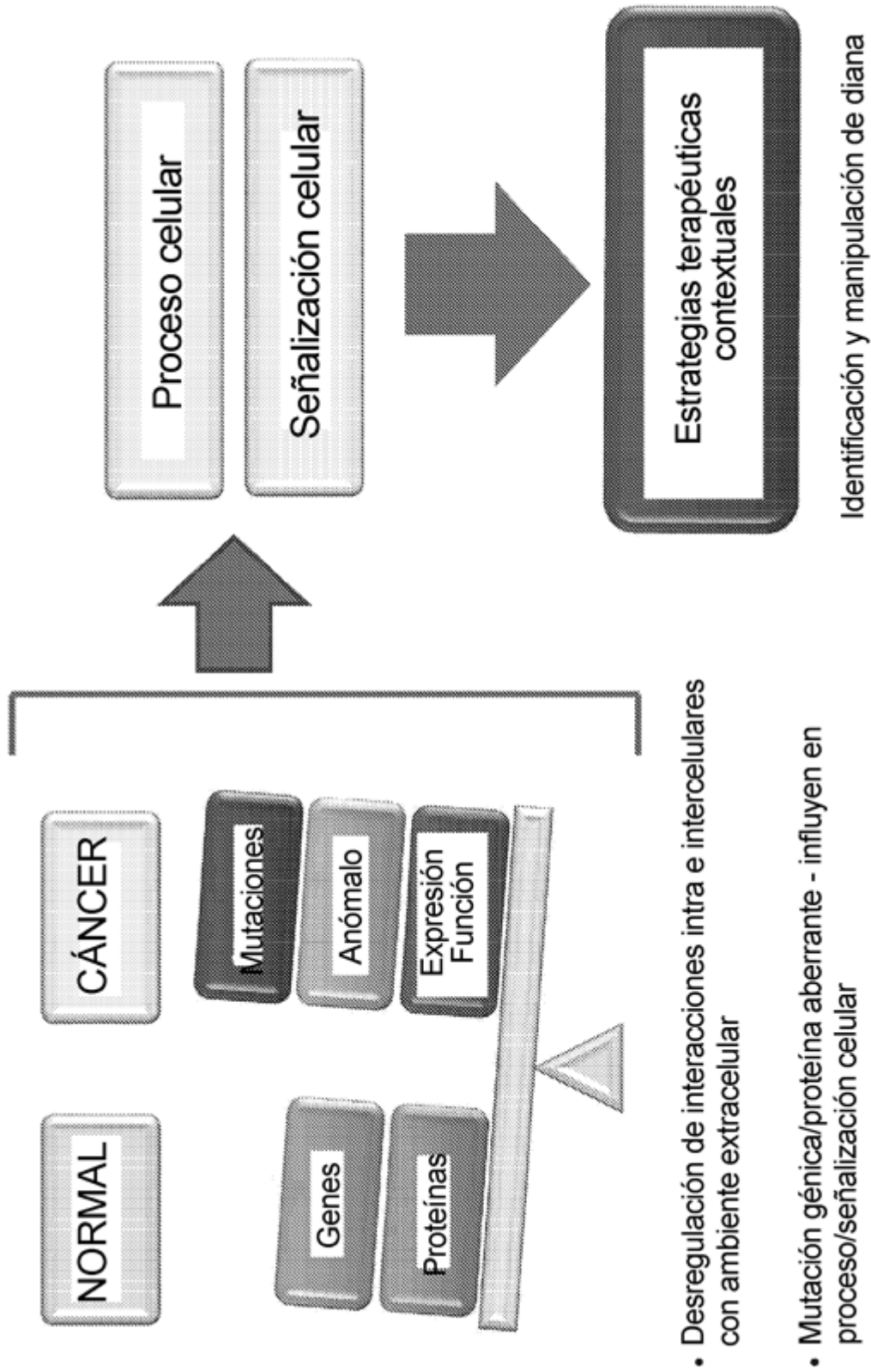


Figura 1

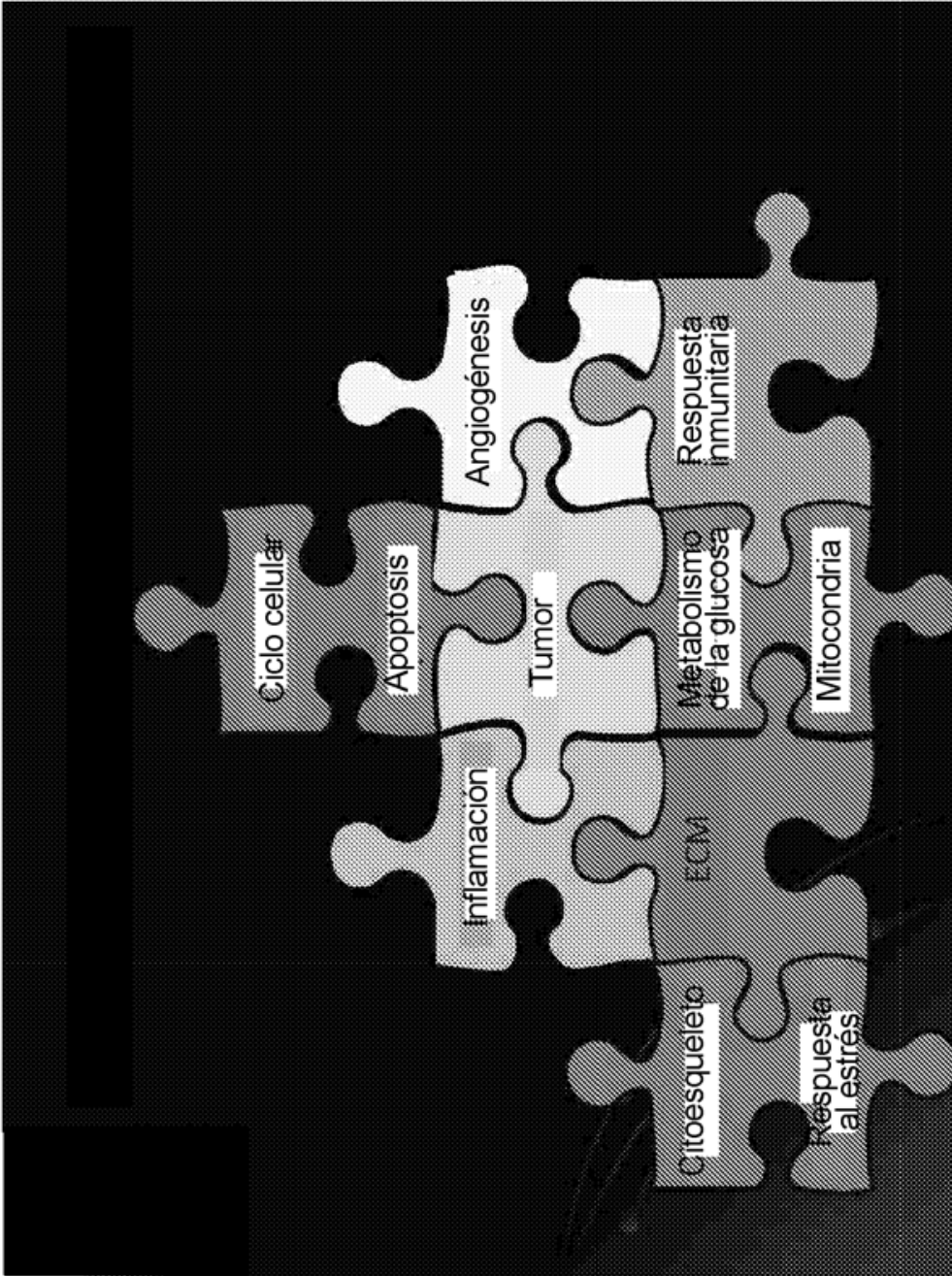


Figura 2

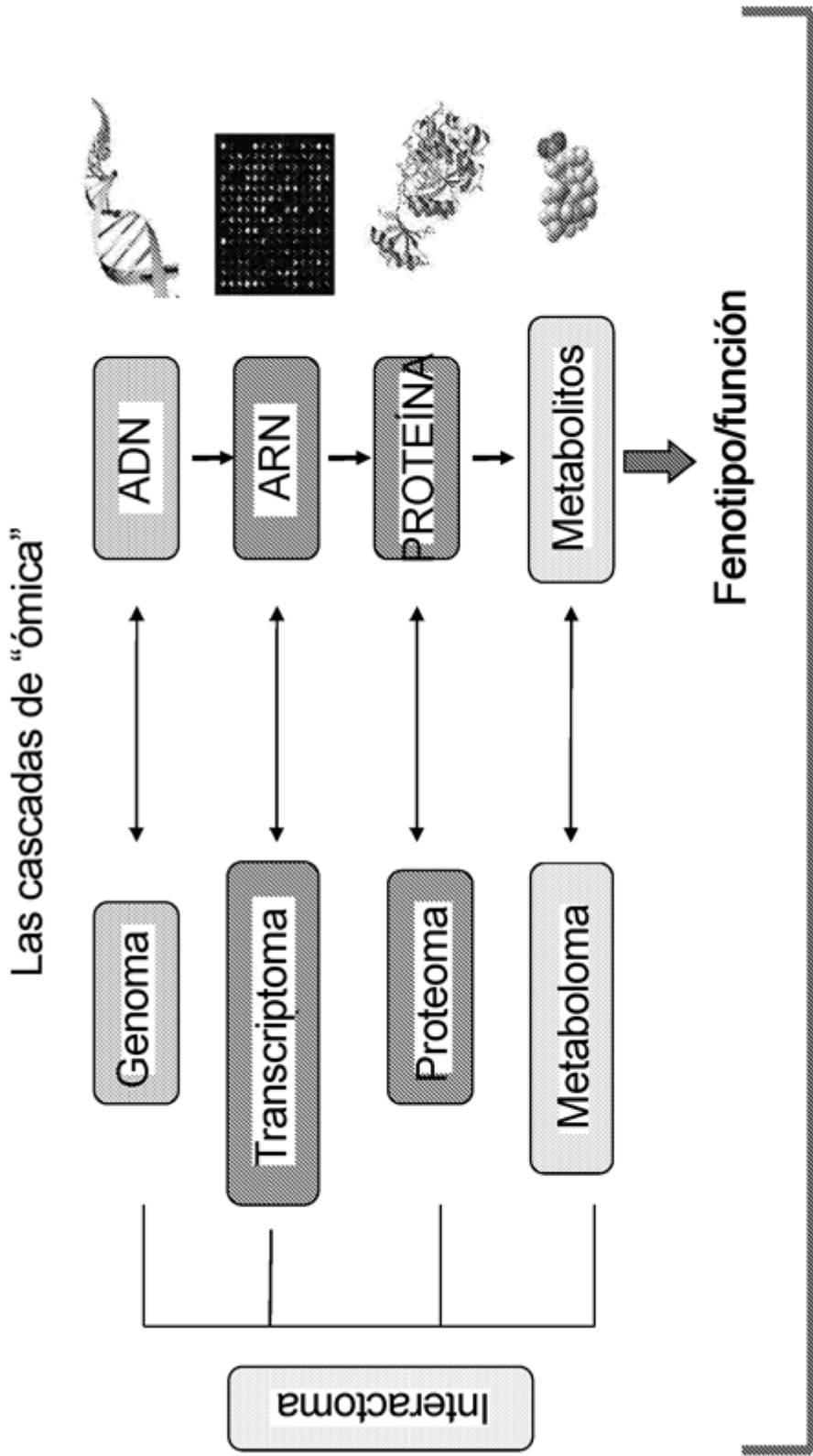


Figura 3

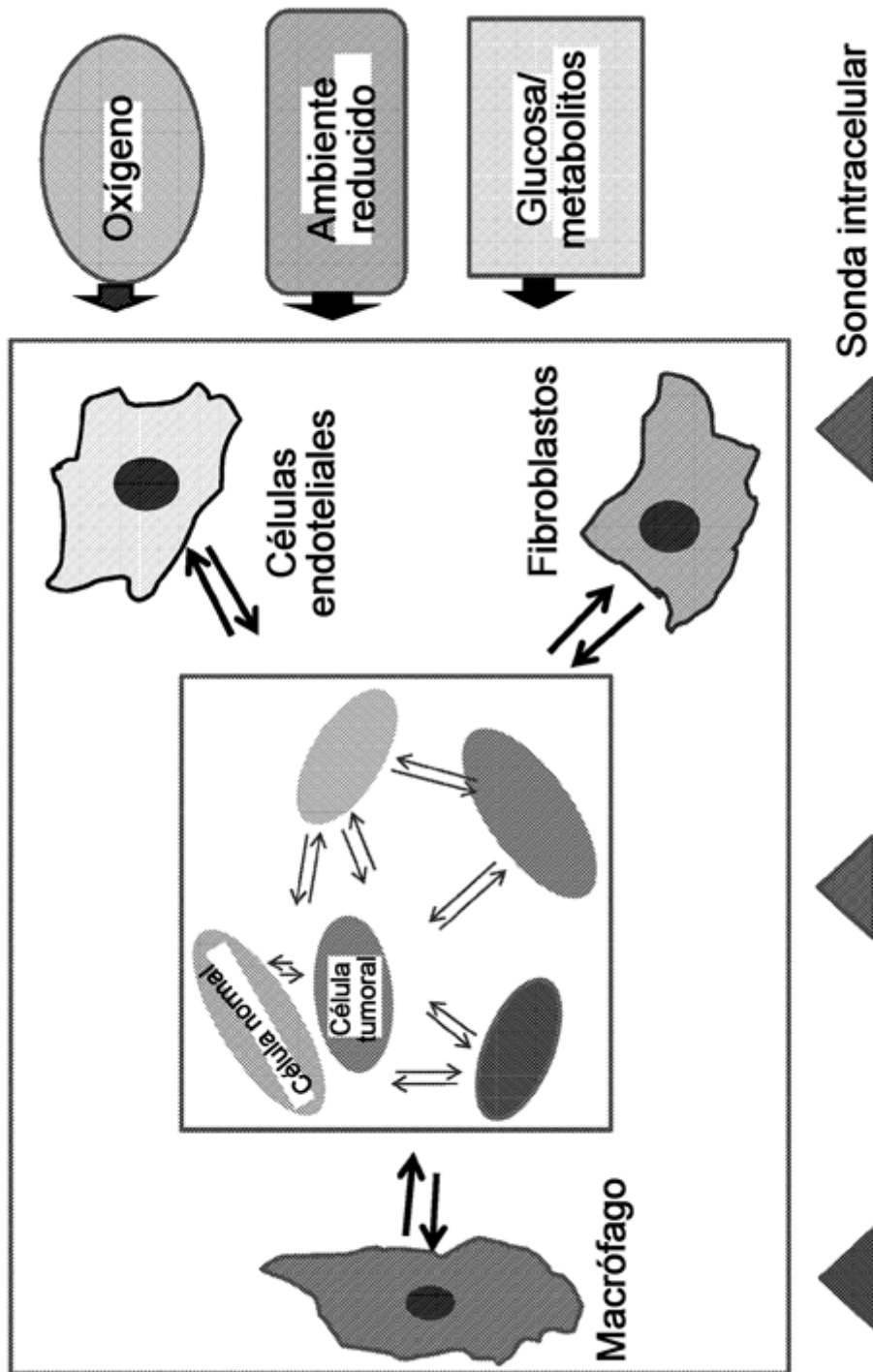


Figura 4

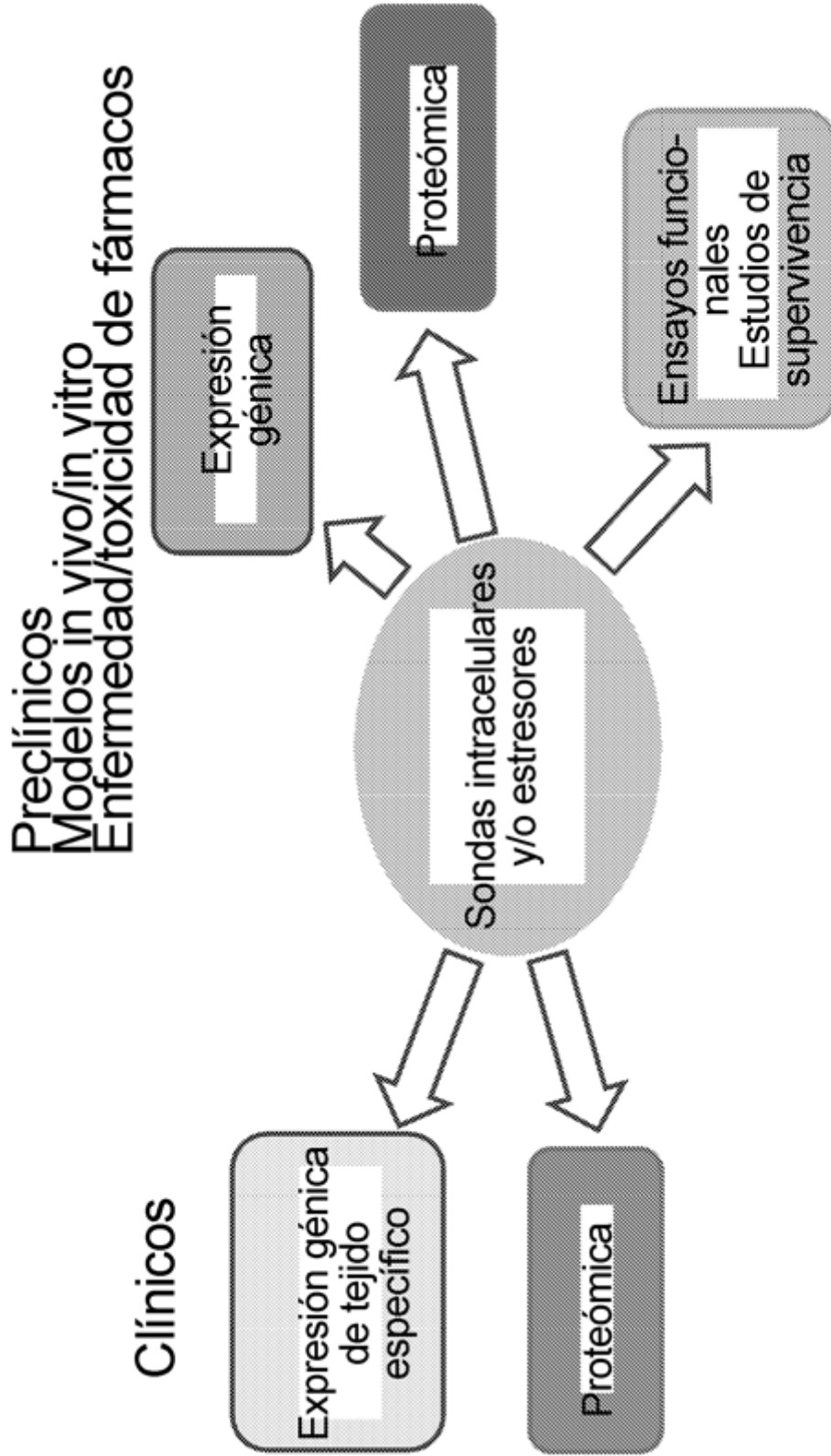


Figura 5

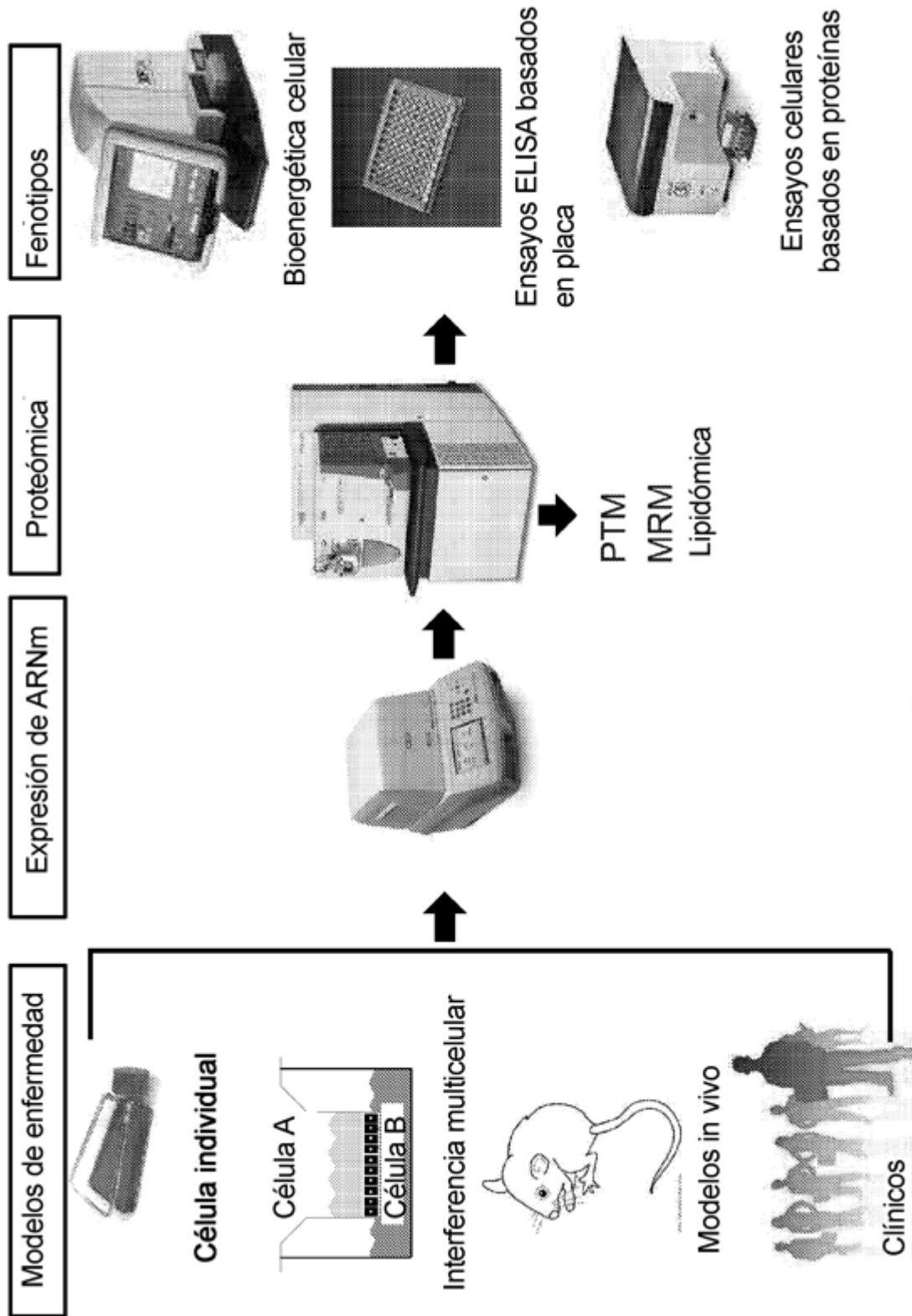


Figura 6

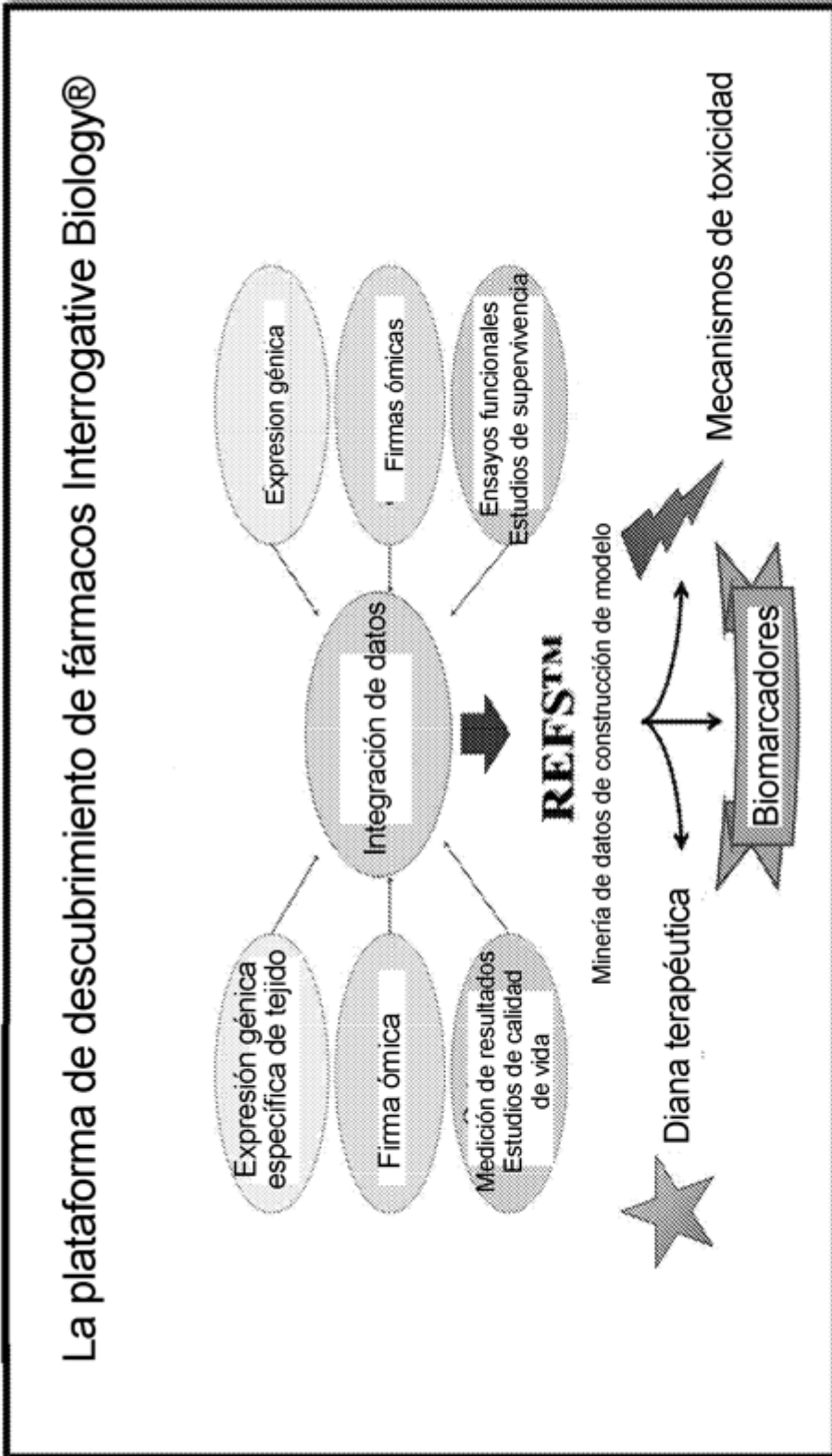


Figura 7

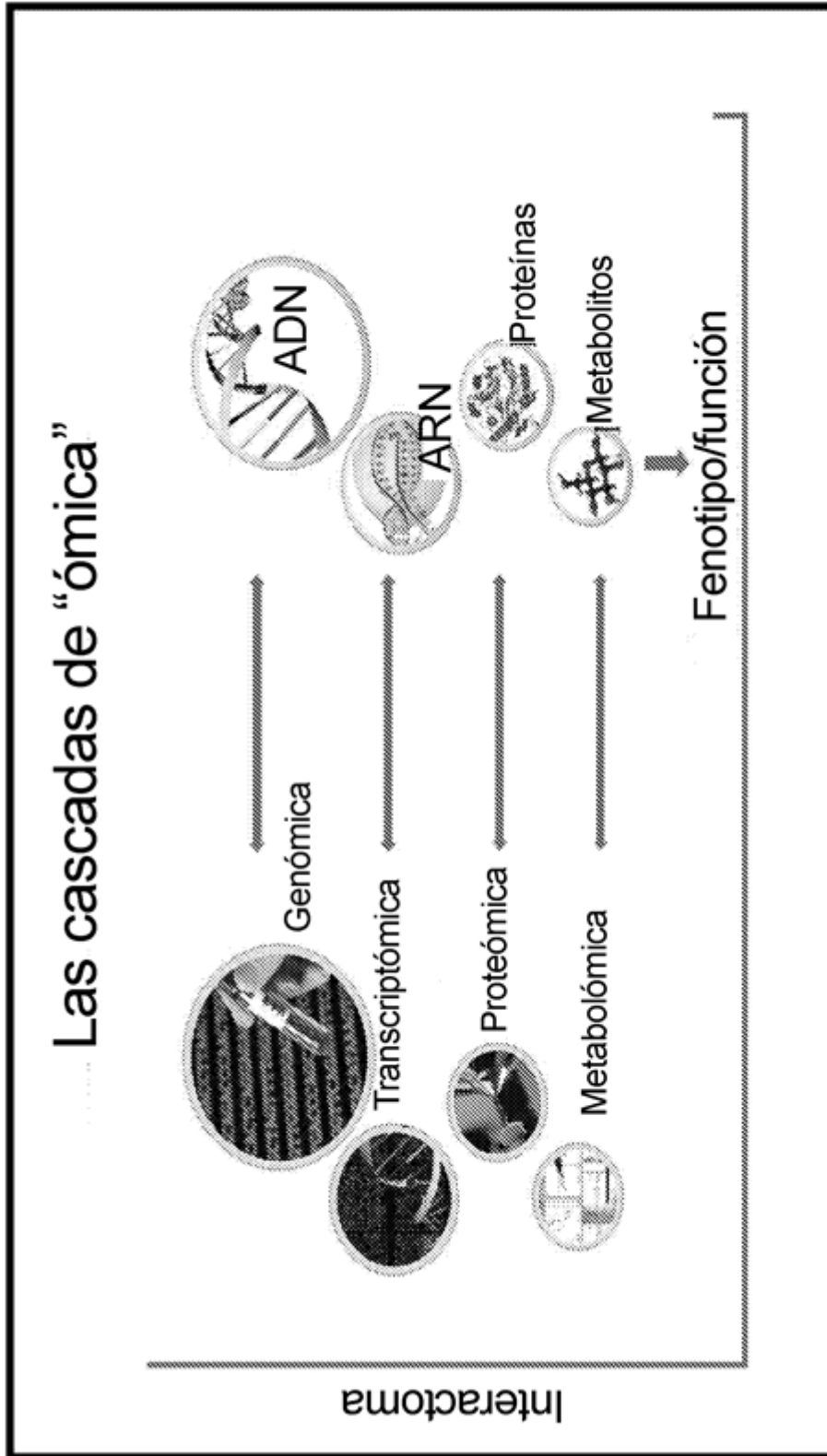


Figura 8

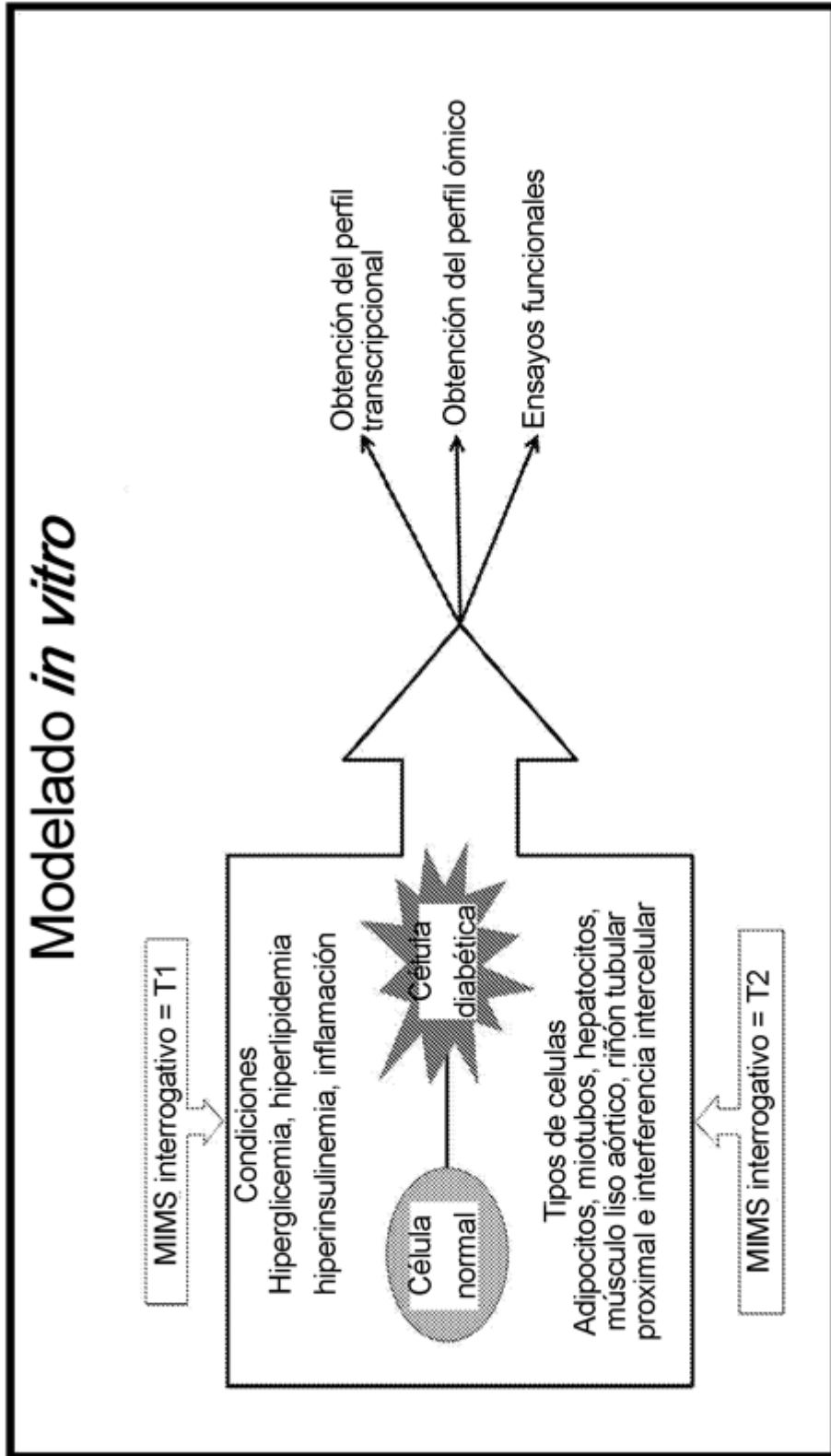


Figura 9

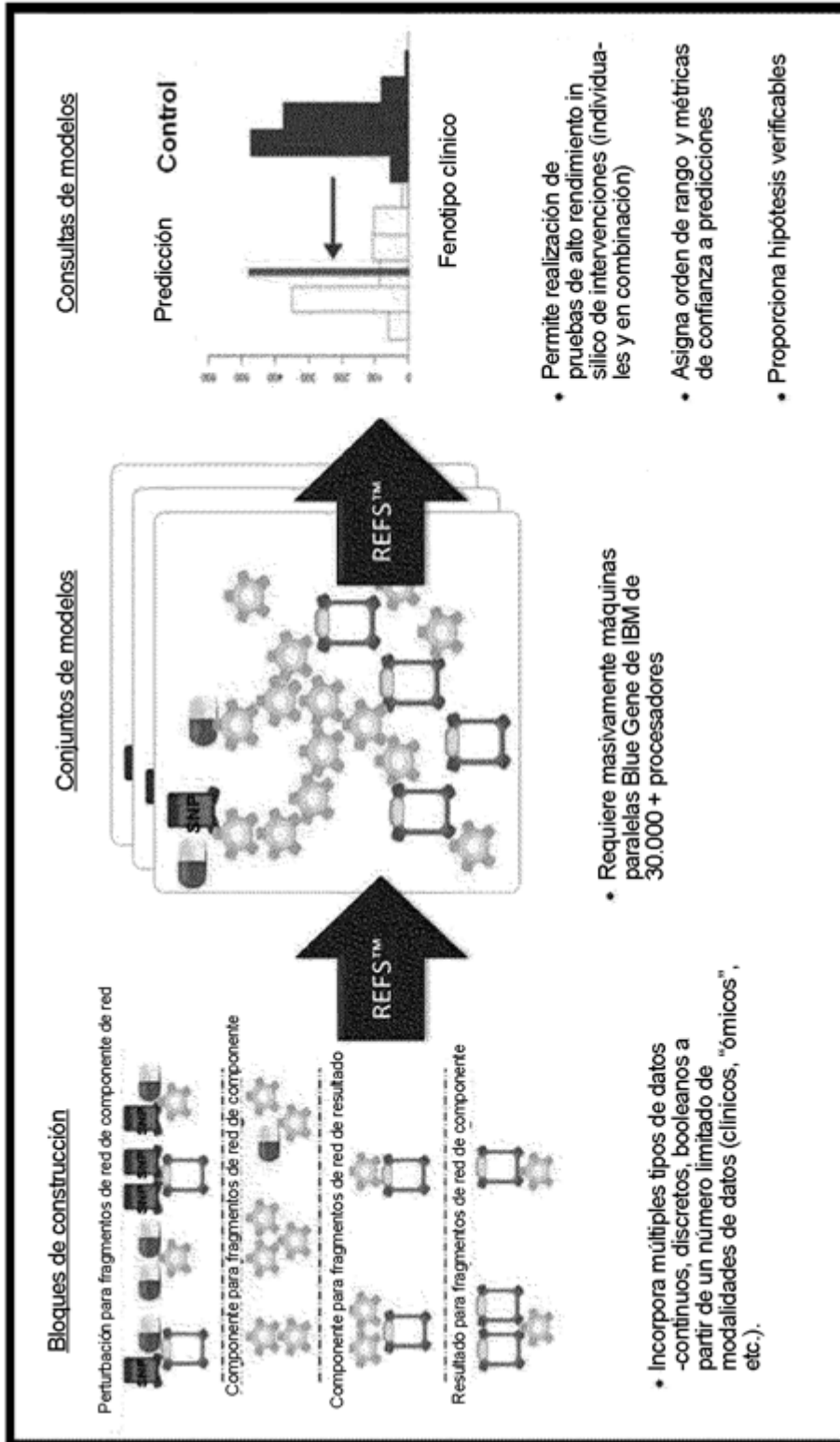


Figura 10

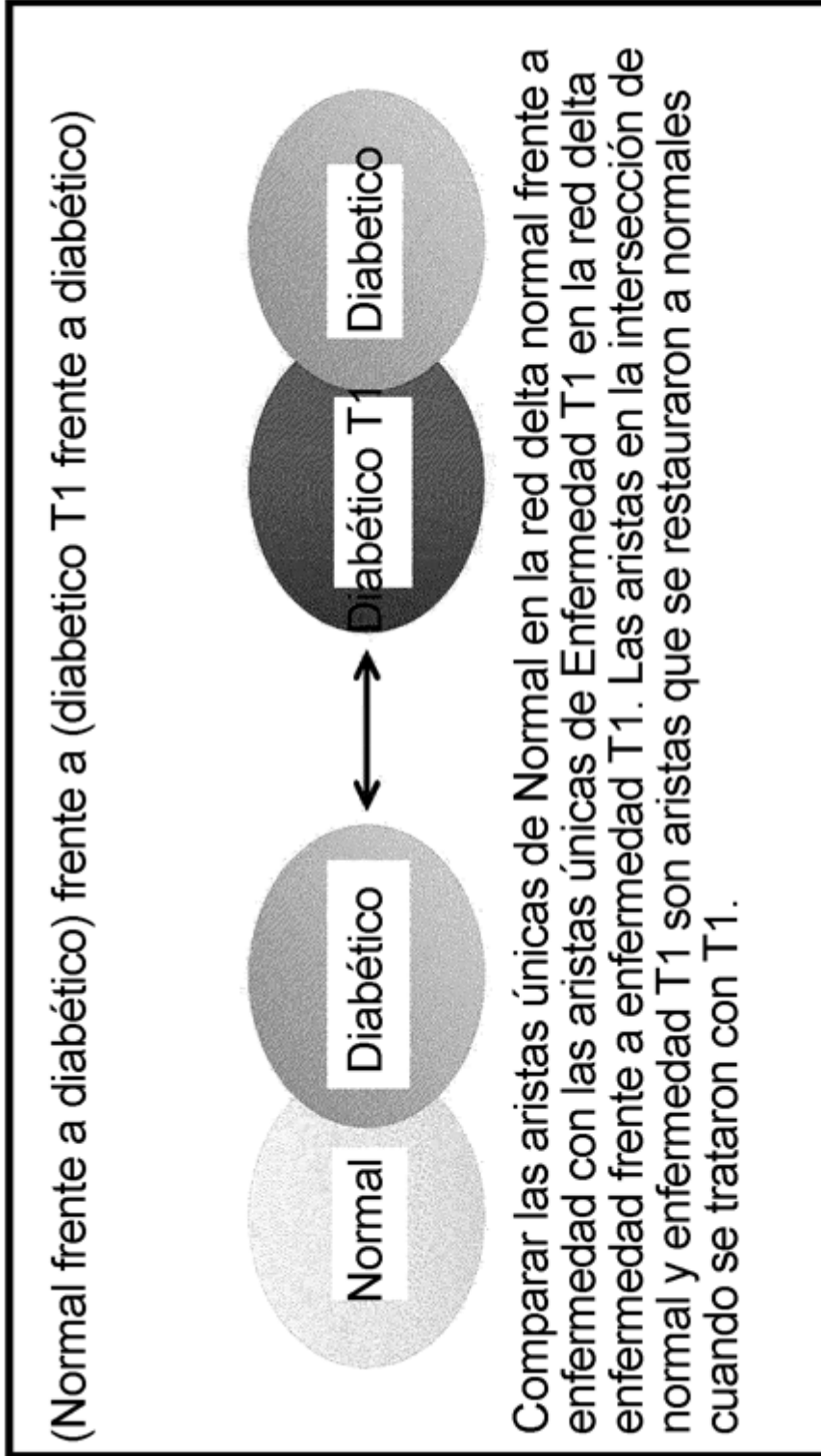


Figura 11

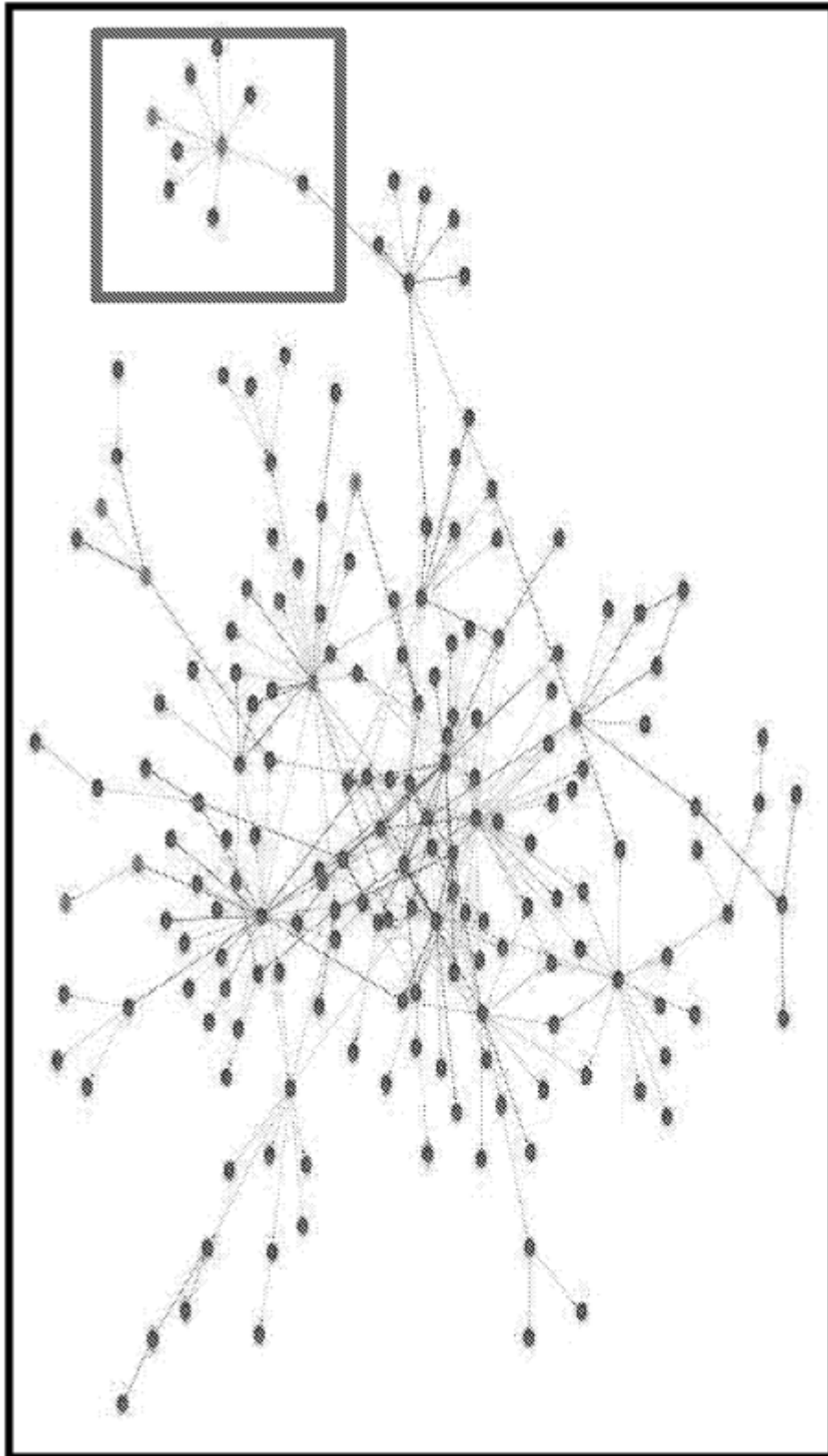


Figura 12

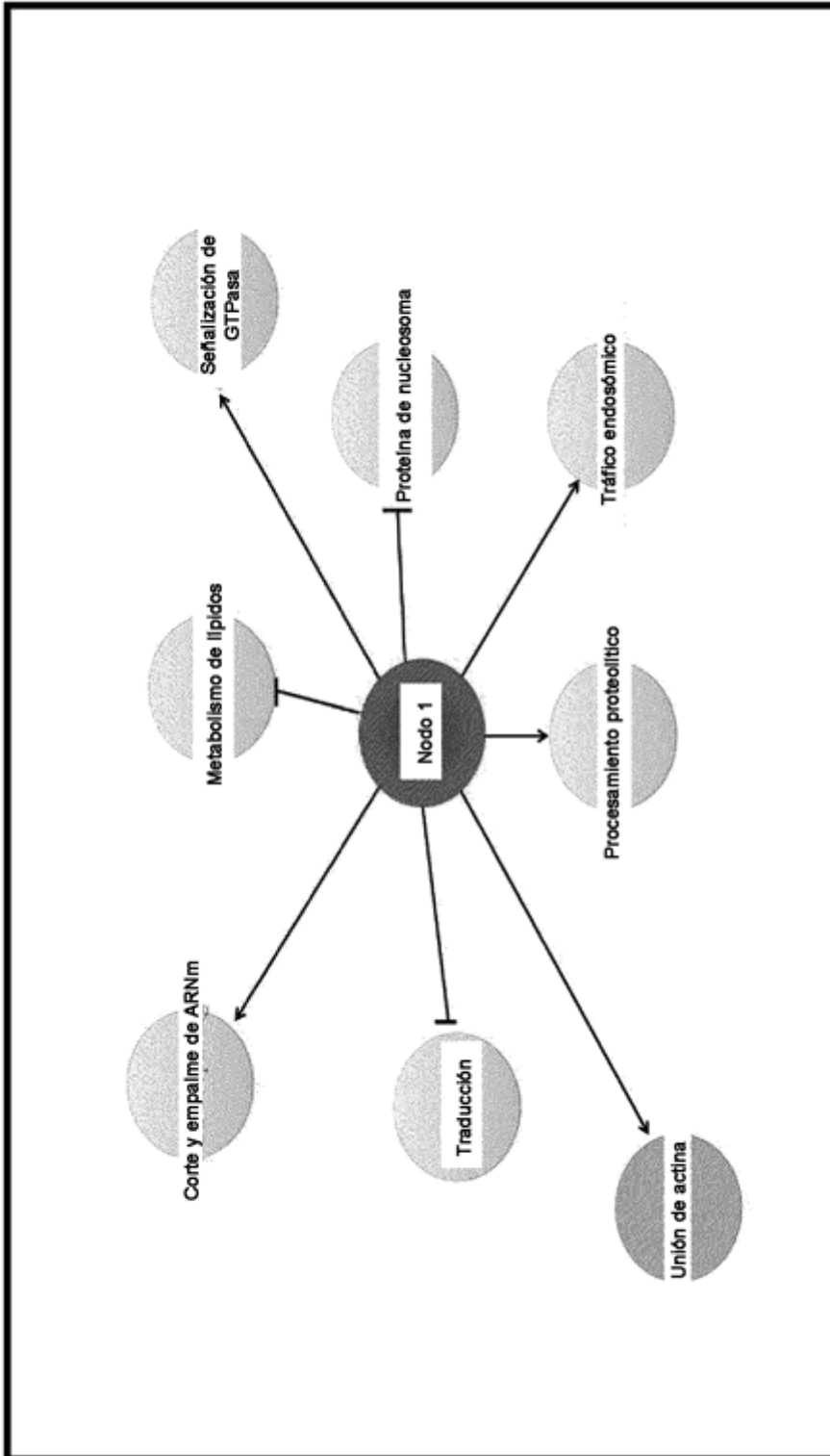


Figura 13

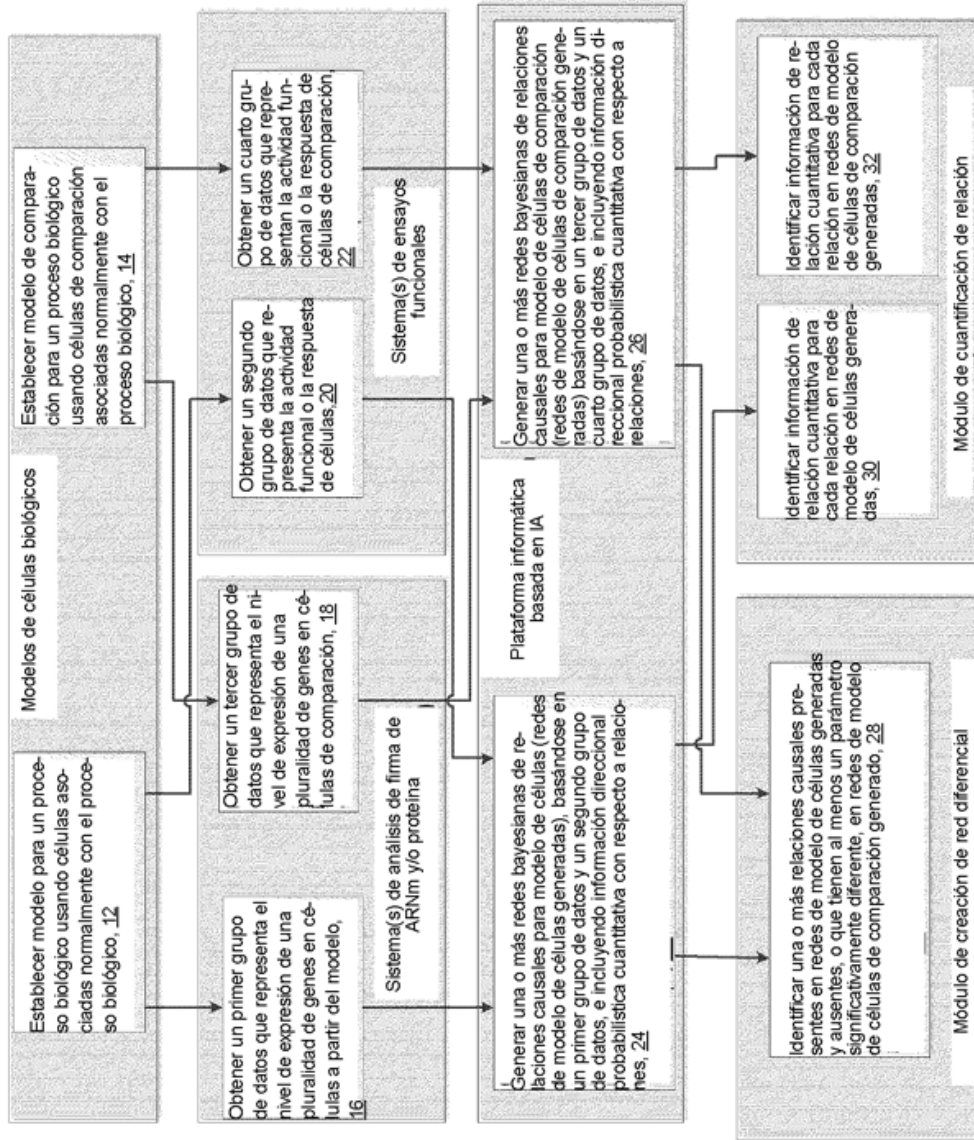


Figura 14

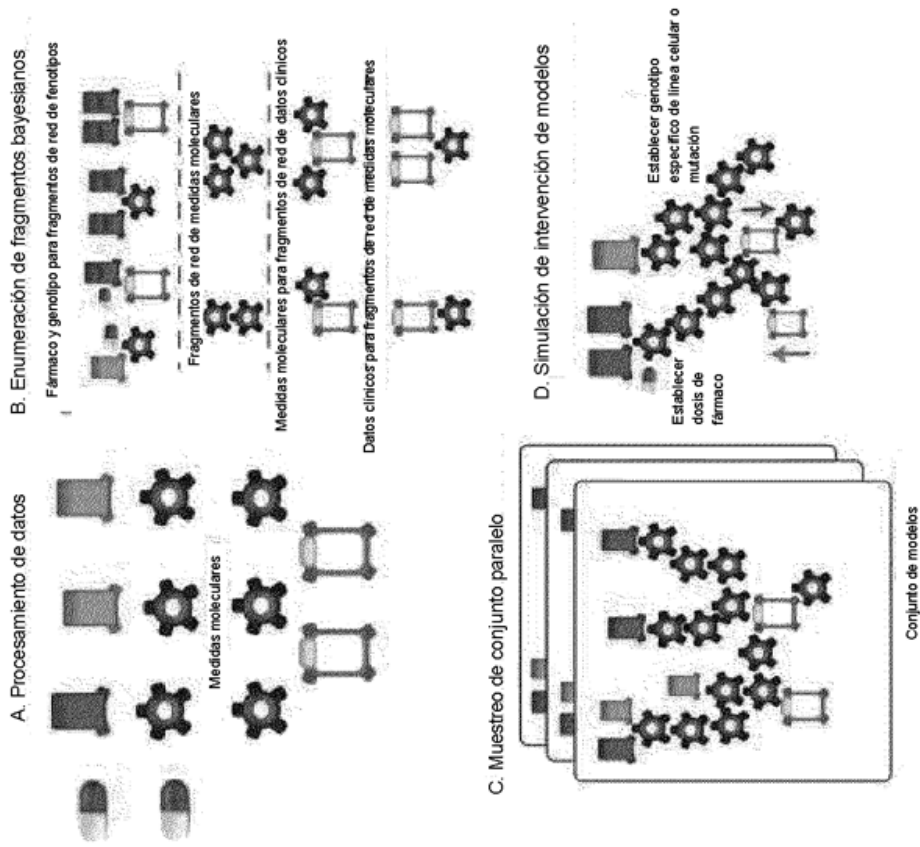


Figura 15

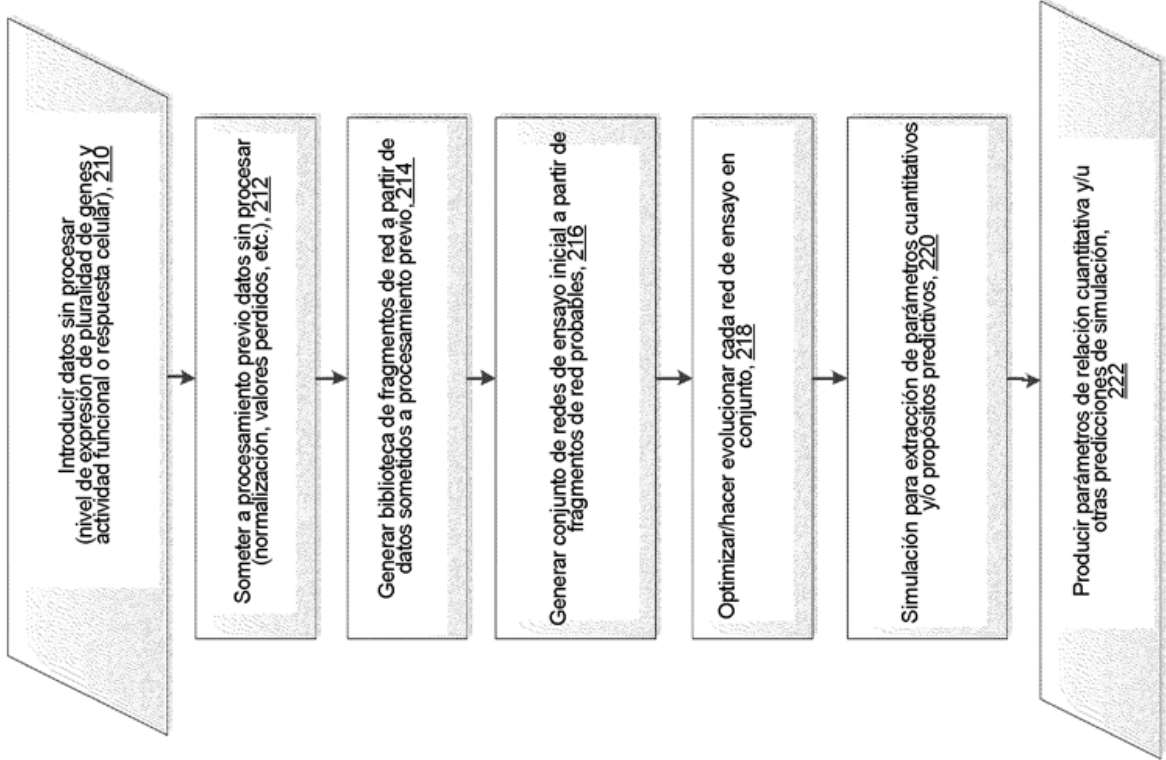


Figura 16

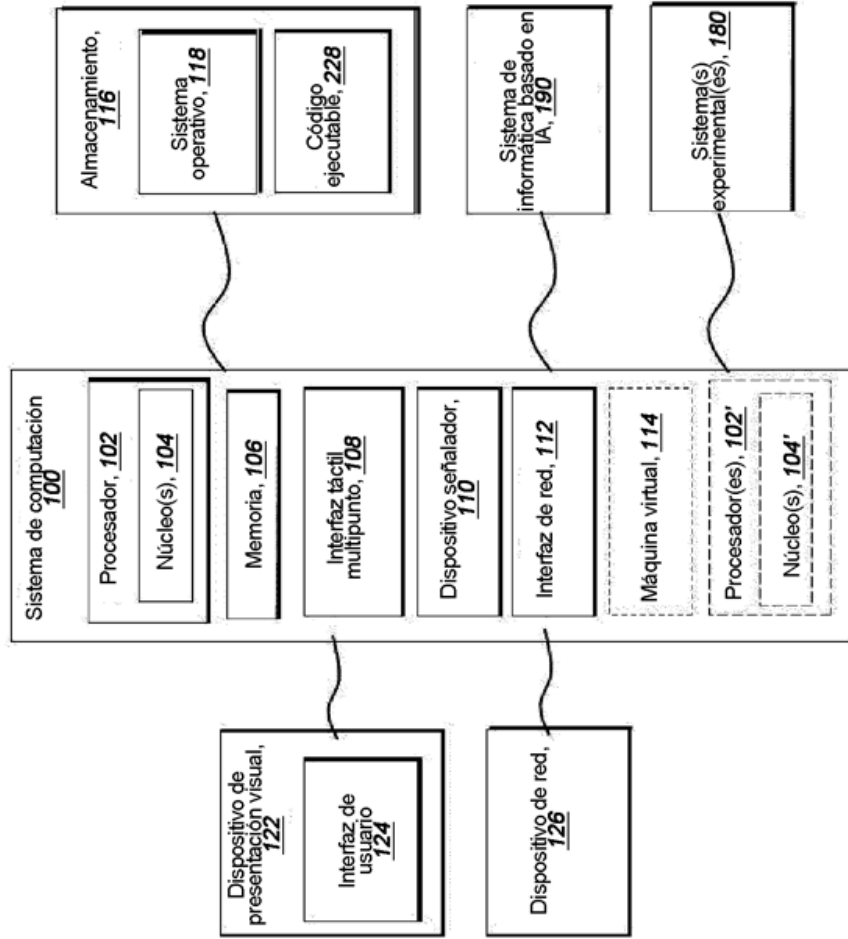


Figura 17

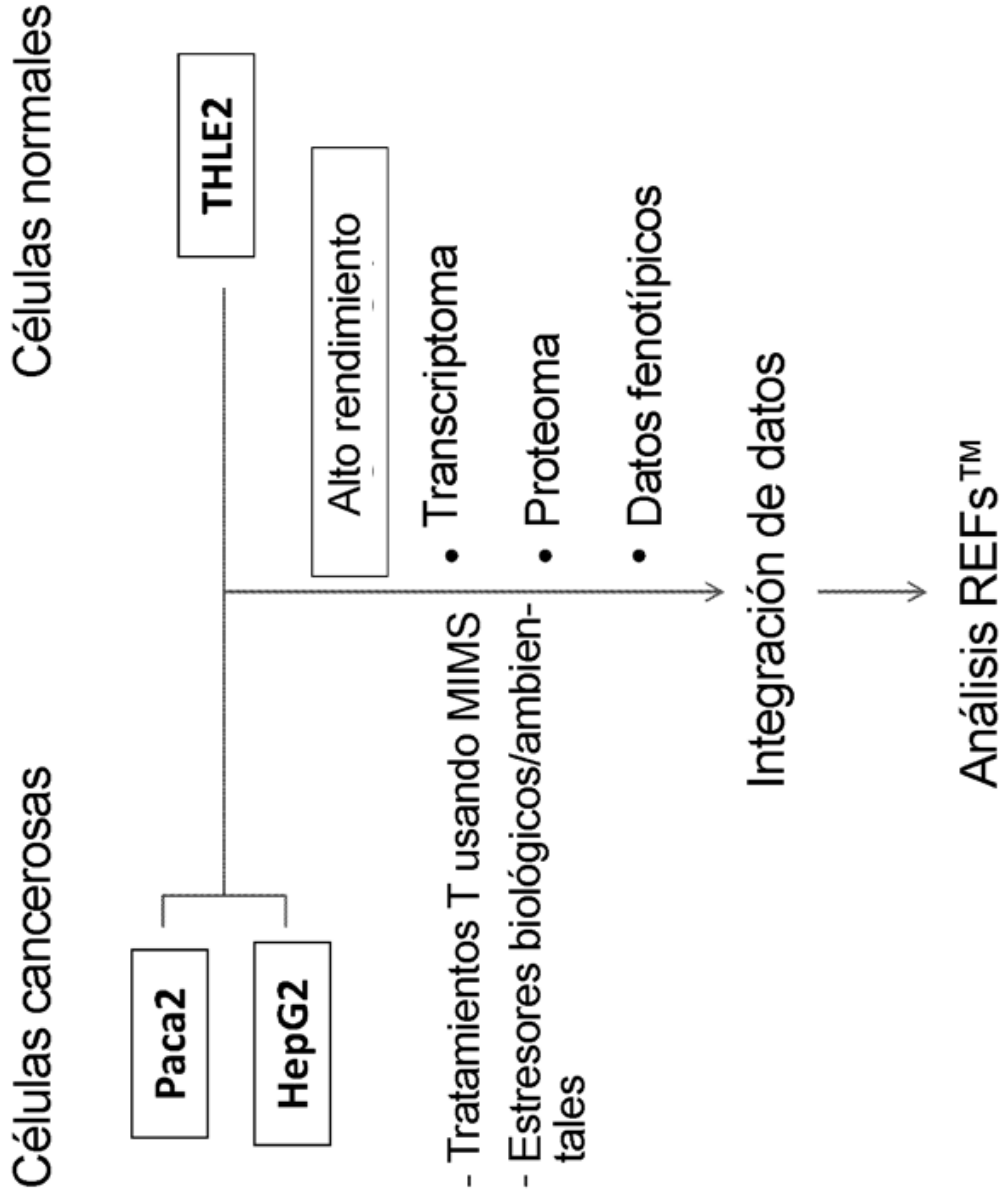
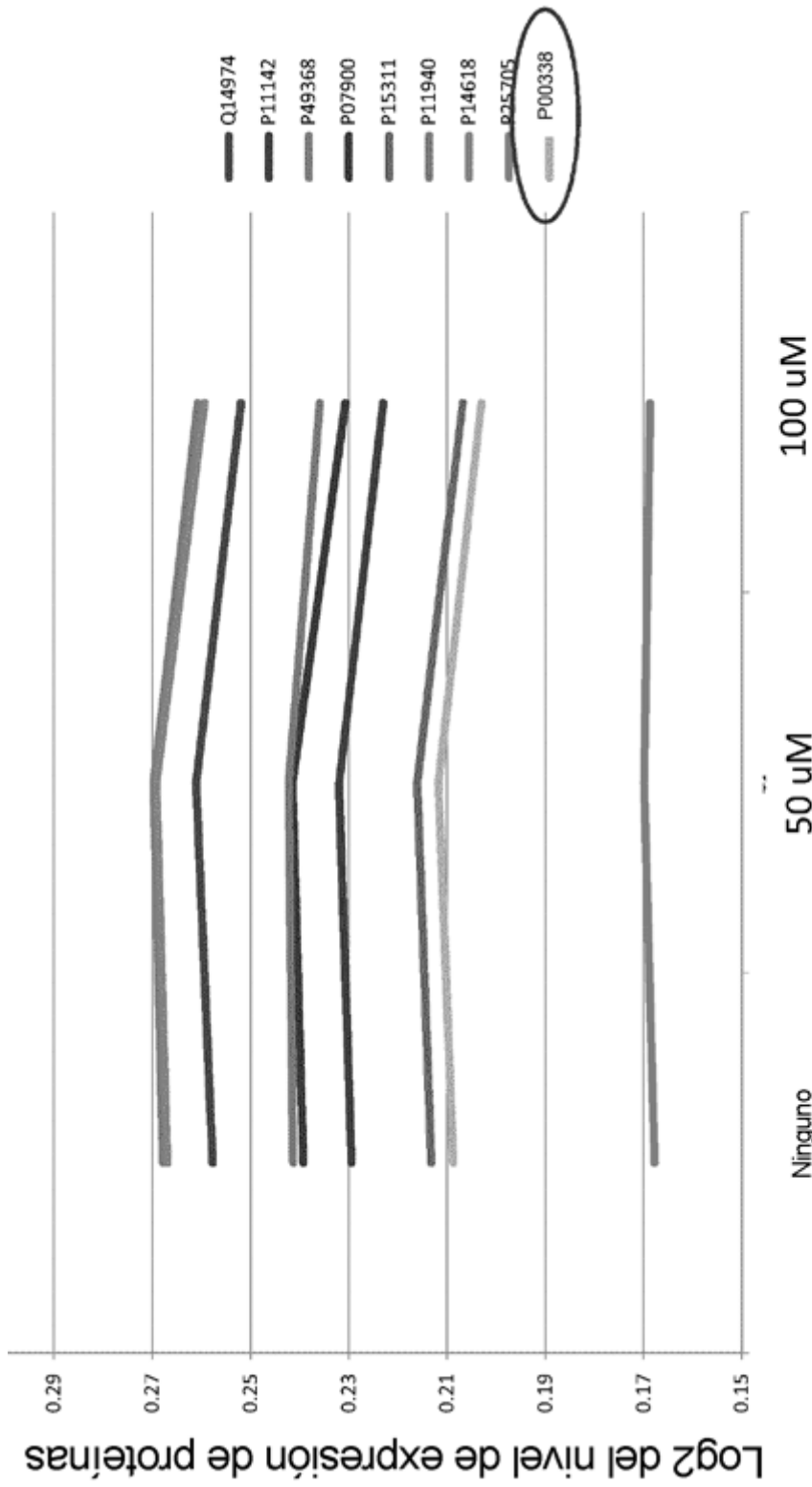


Figura 18



Tratamiento con CoQ10

Figura 19

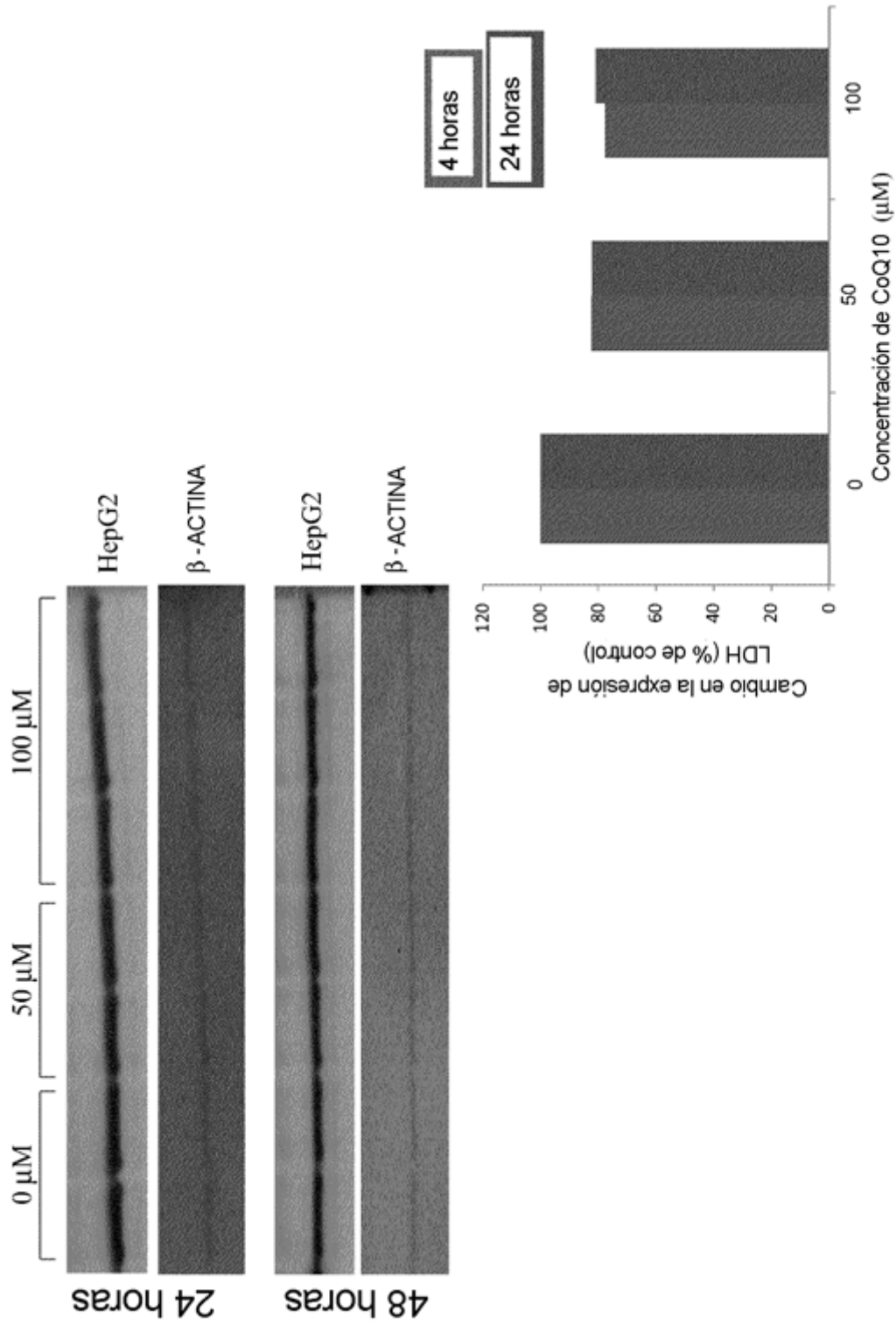


Figura 20

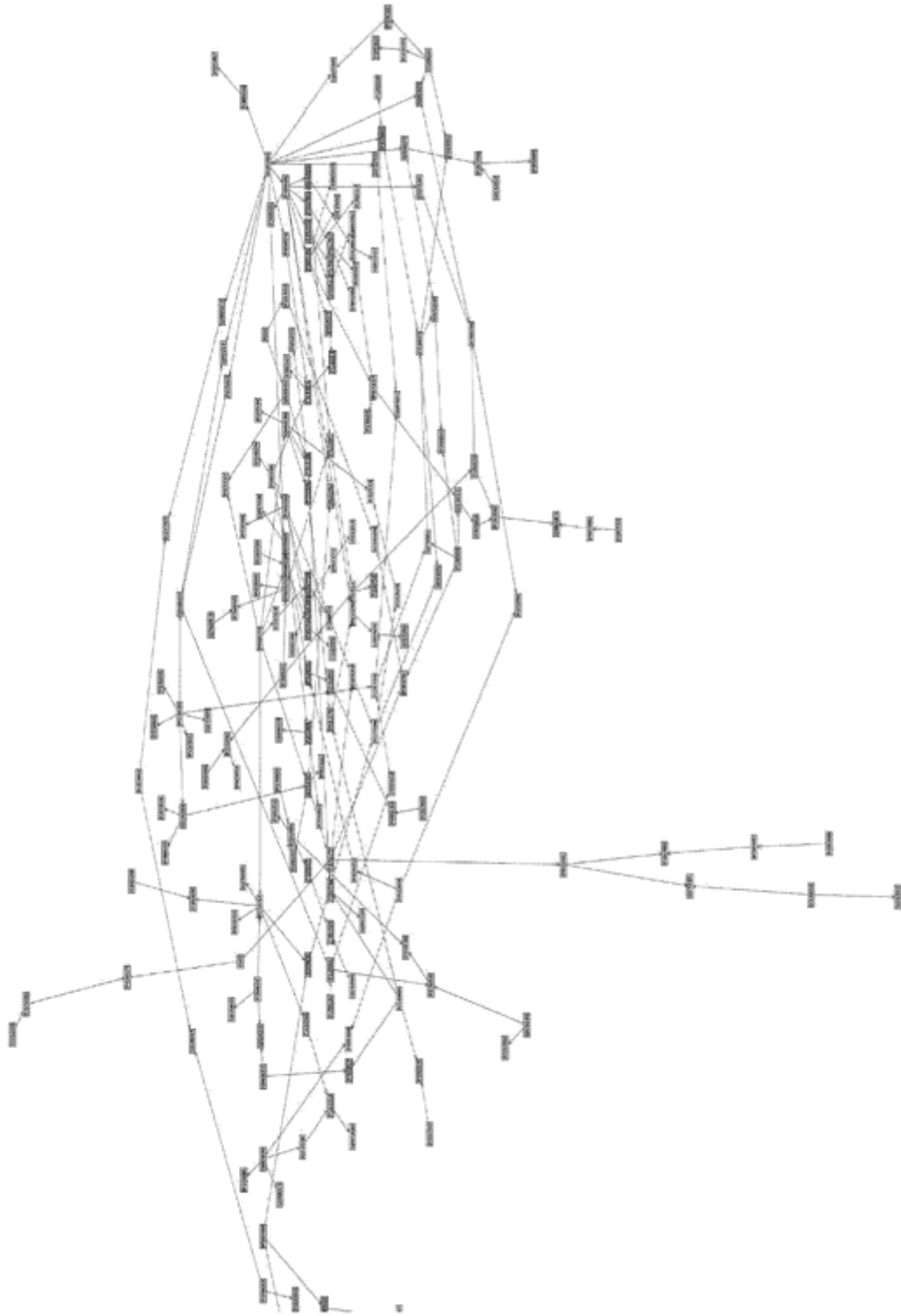


Figura 21

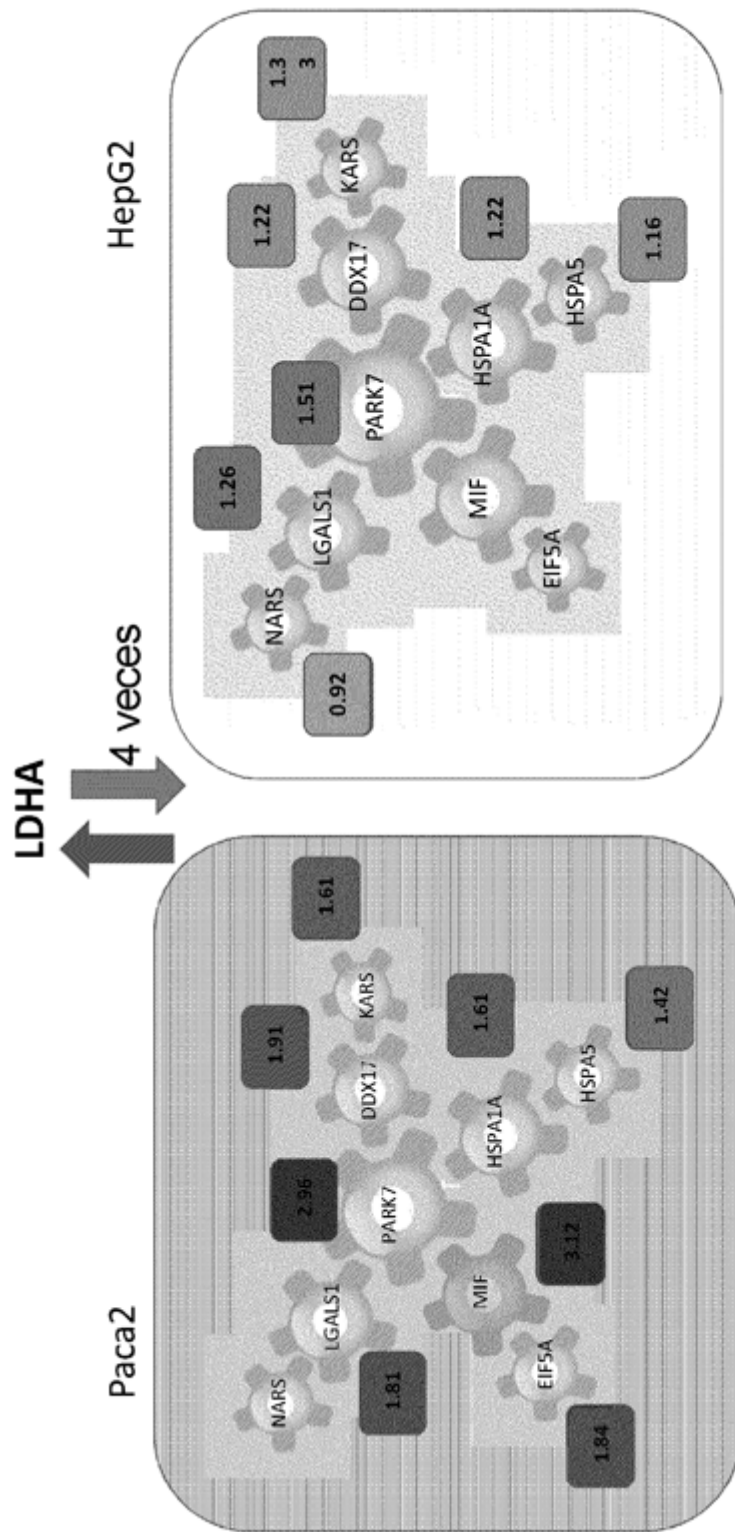


Figura 22

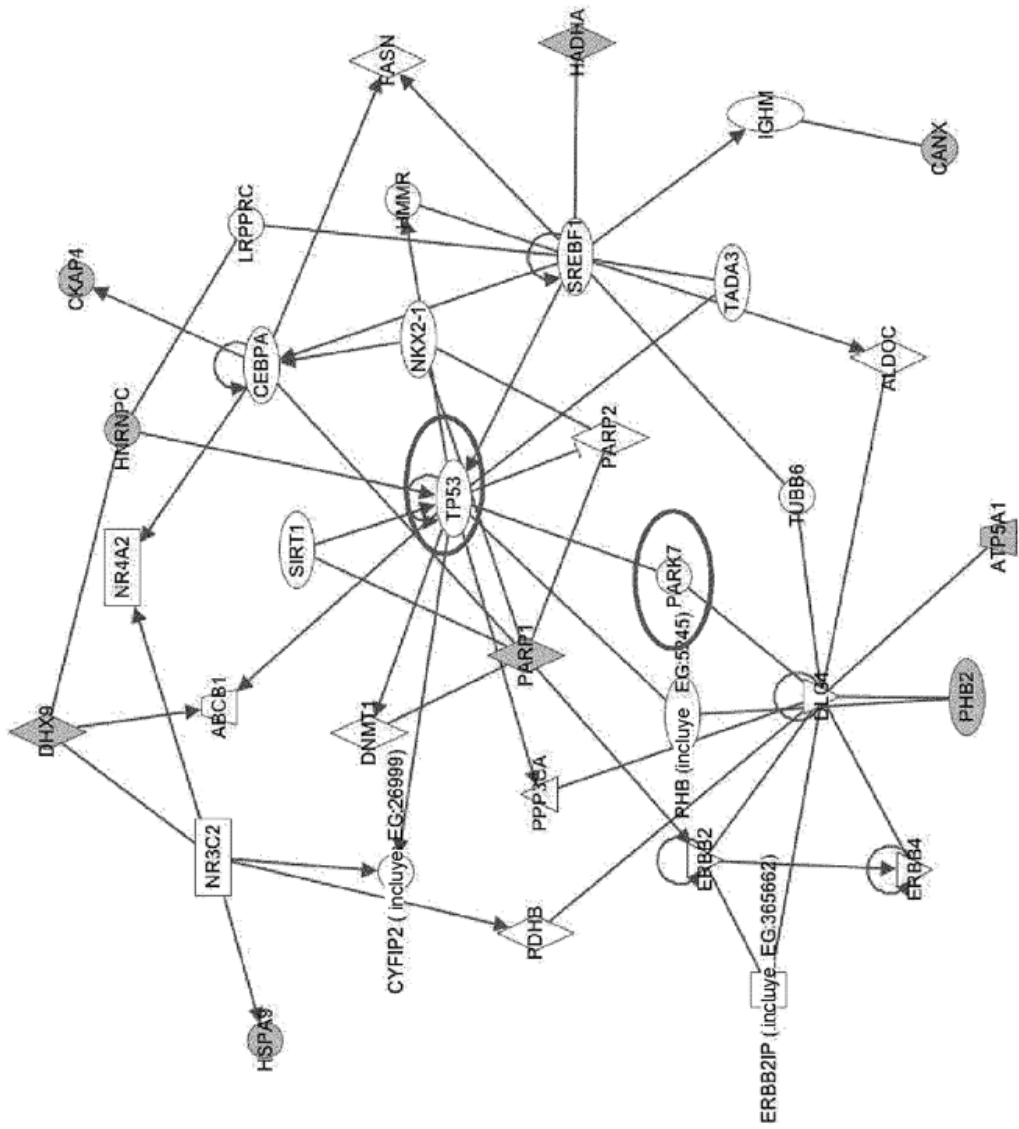


Figura 23

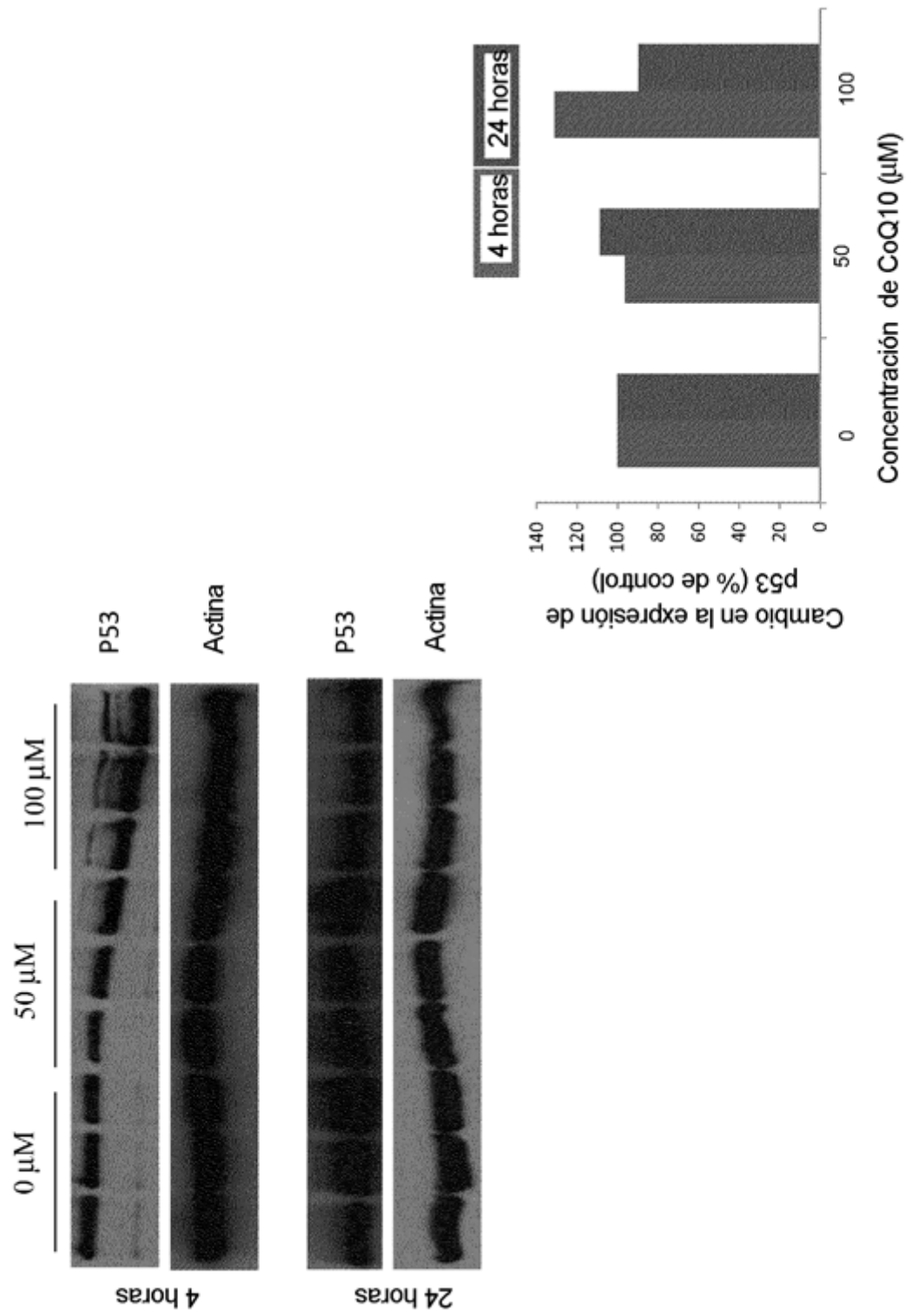


Figura 24

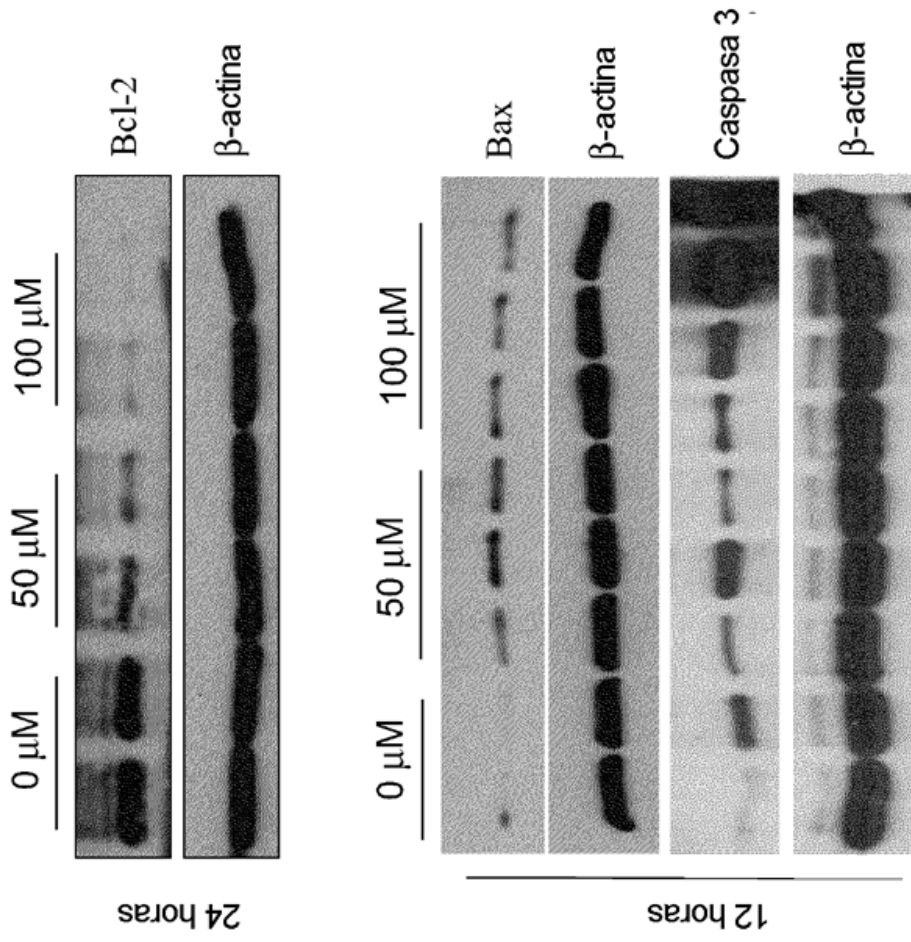


Figura 25

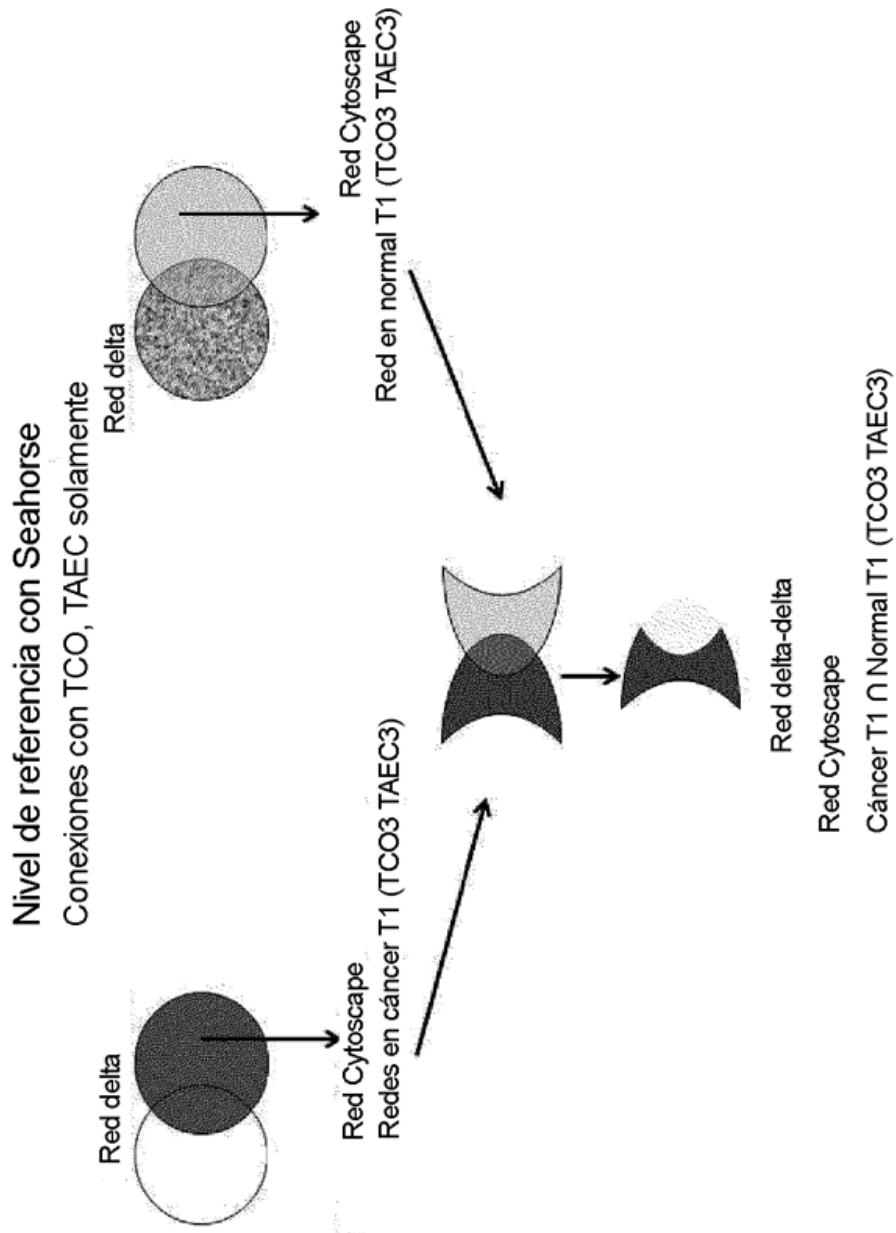


Figura 26

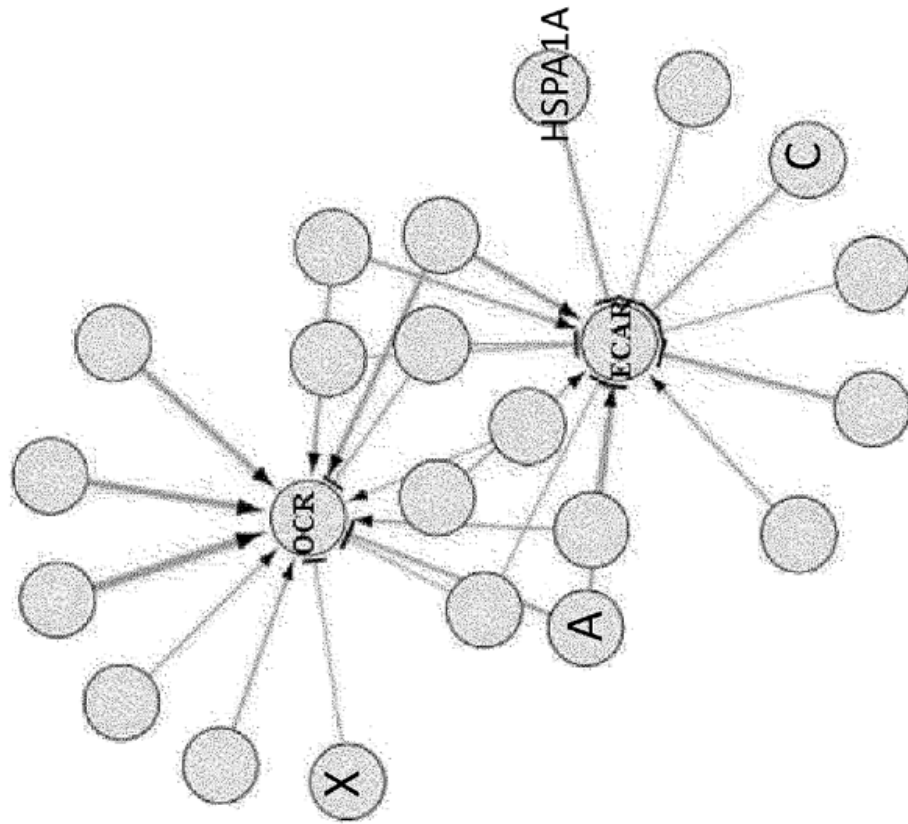


Figura 27

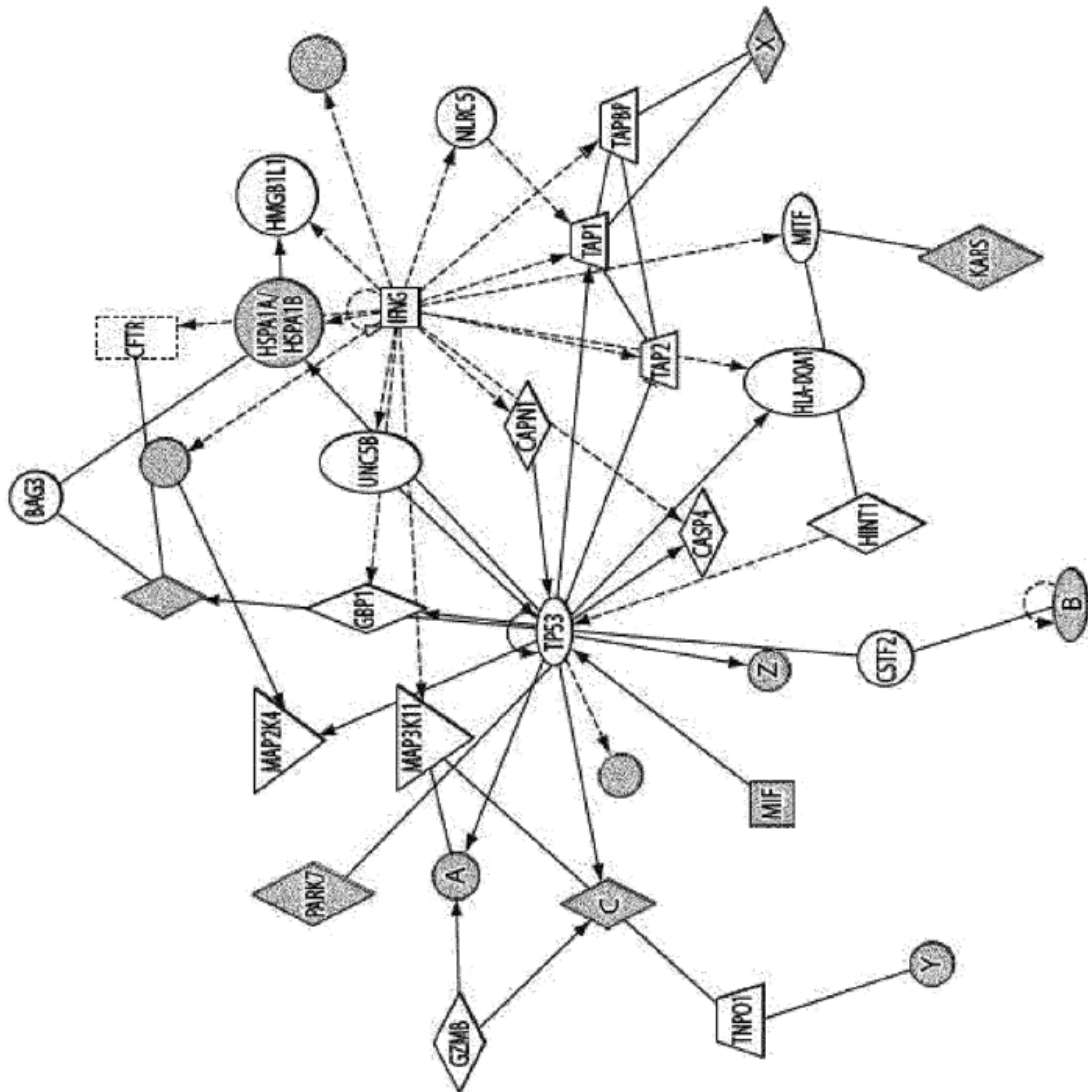


Figure 28

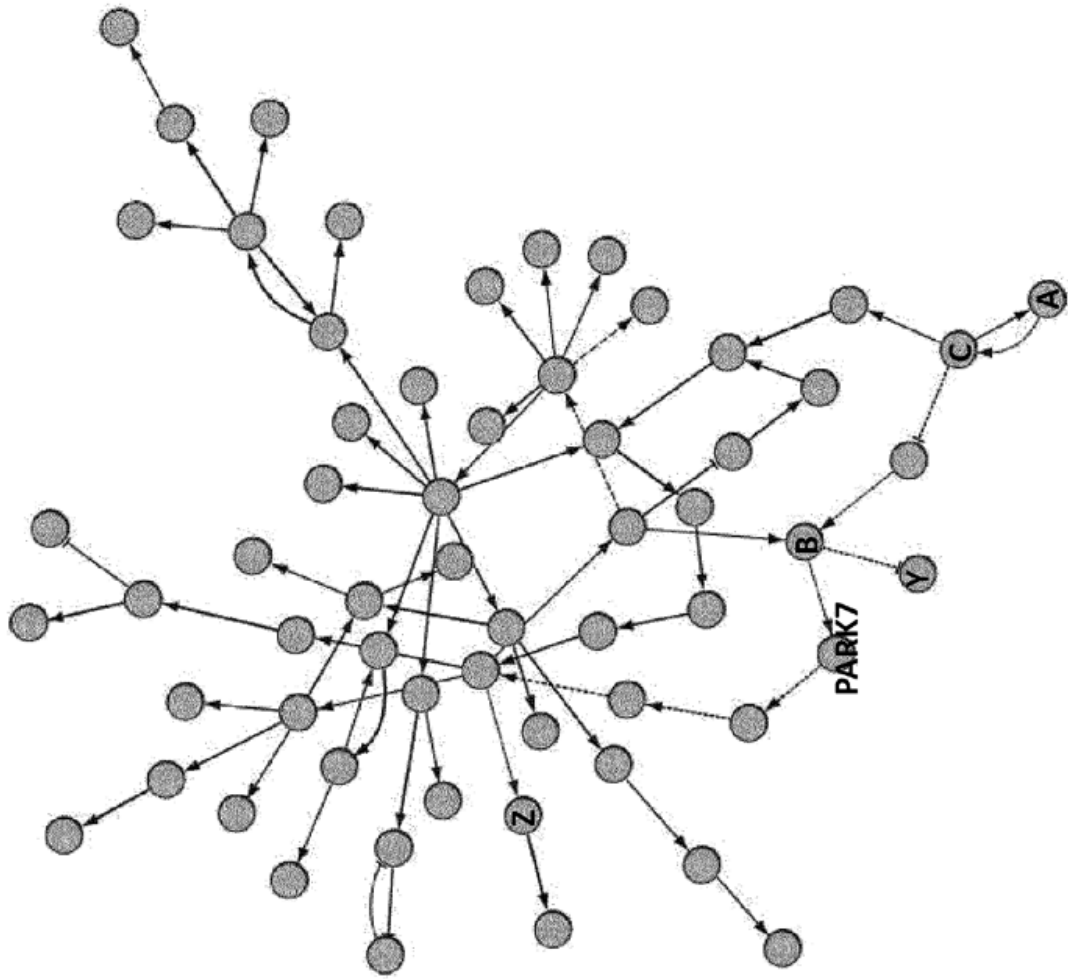


Figura 29

Redes delta-delta

DD_NG.vs.HG.vs.HGT1.cys

[NG frente a HG] frente a [HG frente a HGT1]

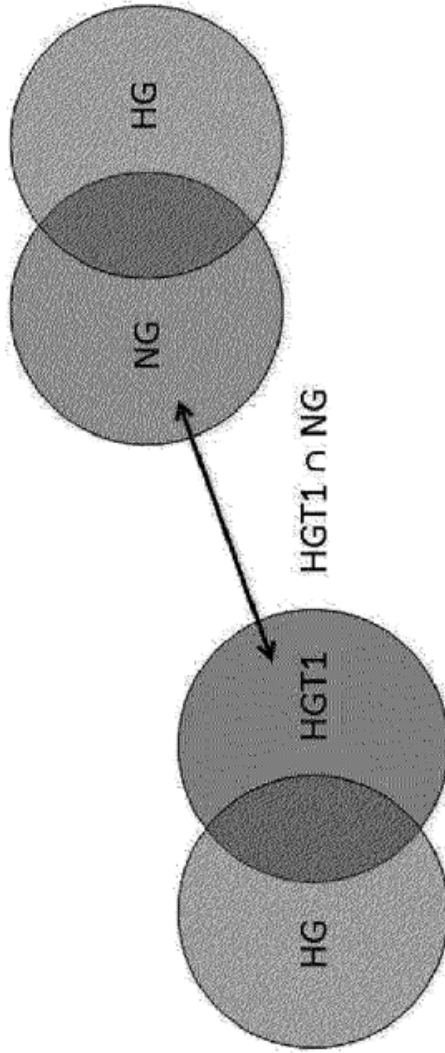


Figura 30

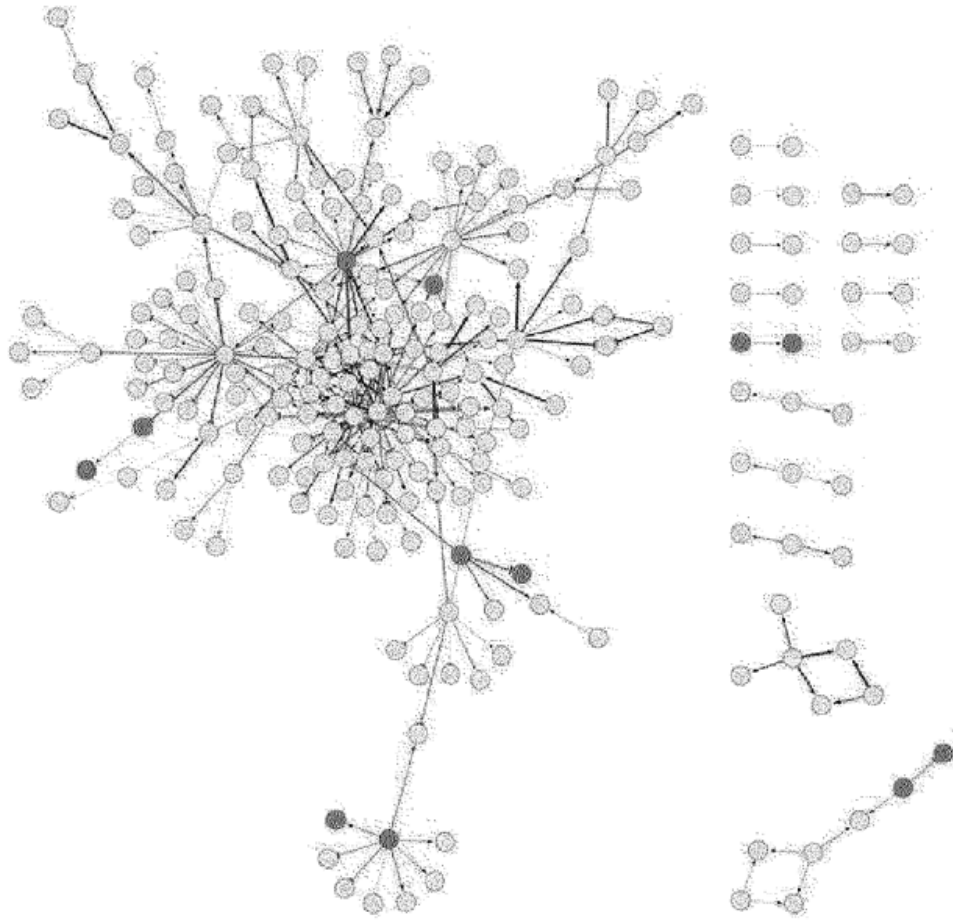


Figura 31

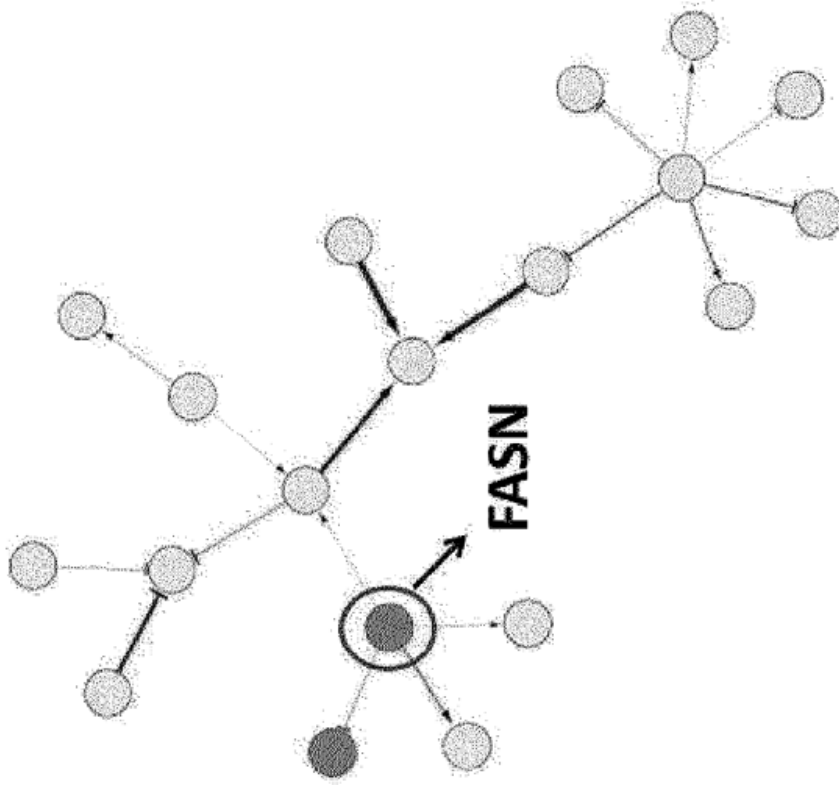


Figura 32

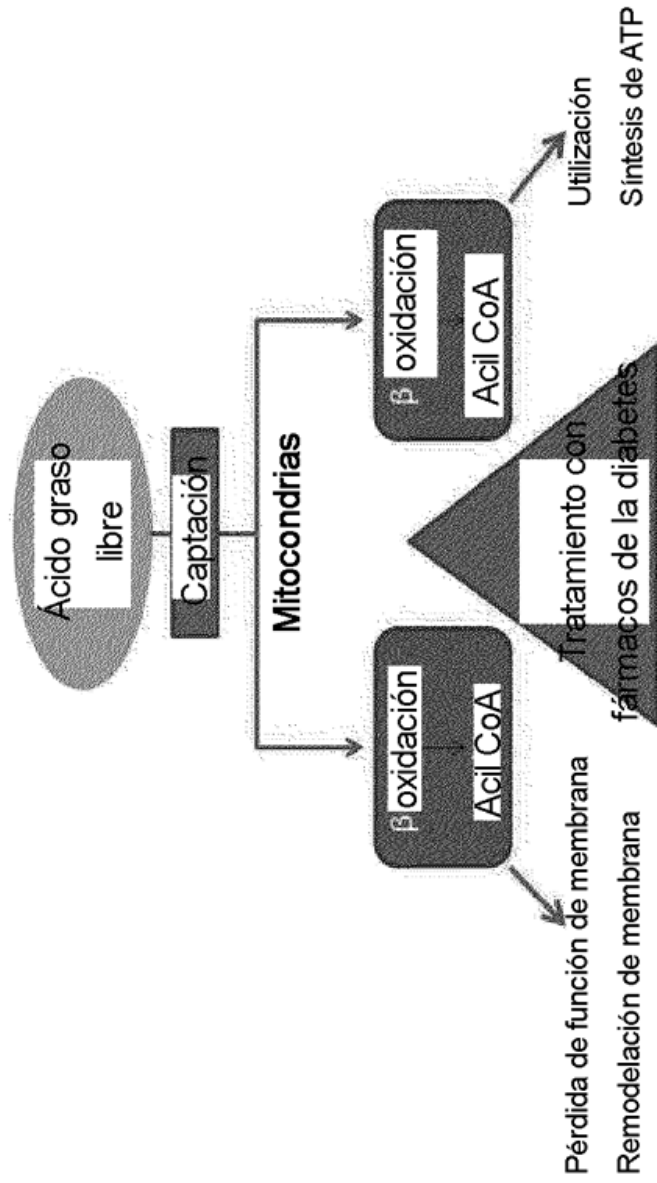


Figura 33

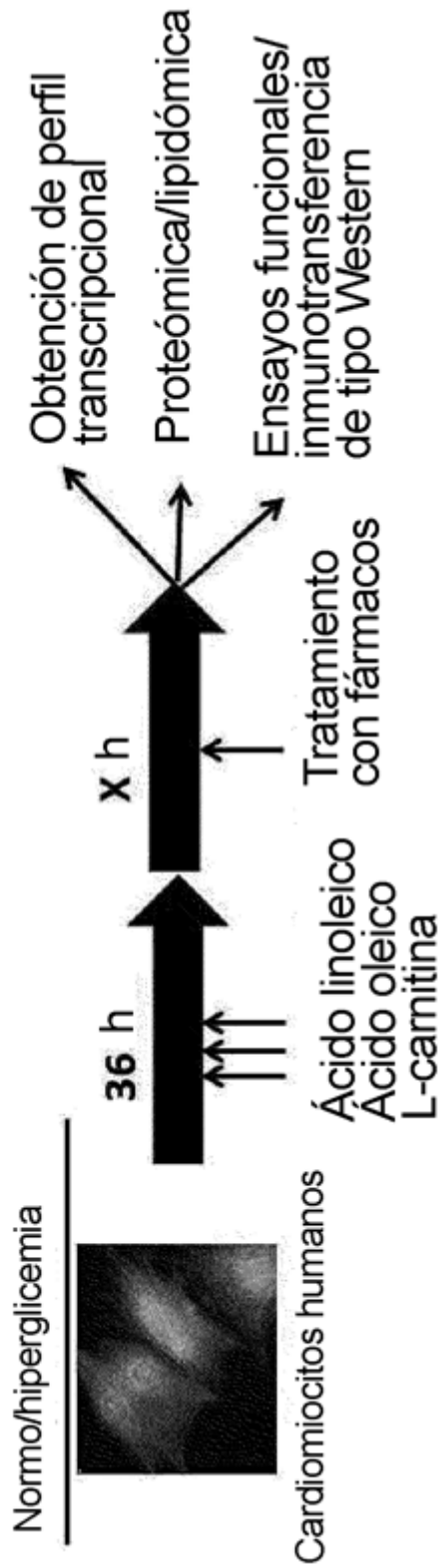


Figura 34

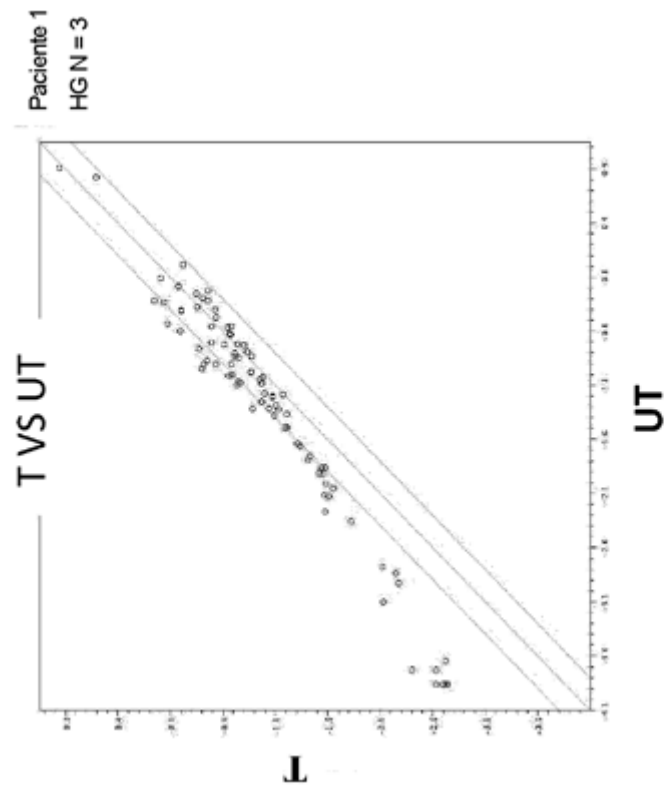
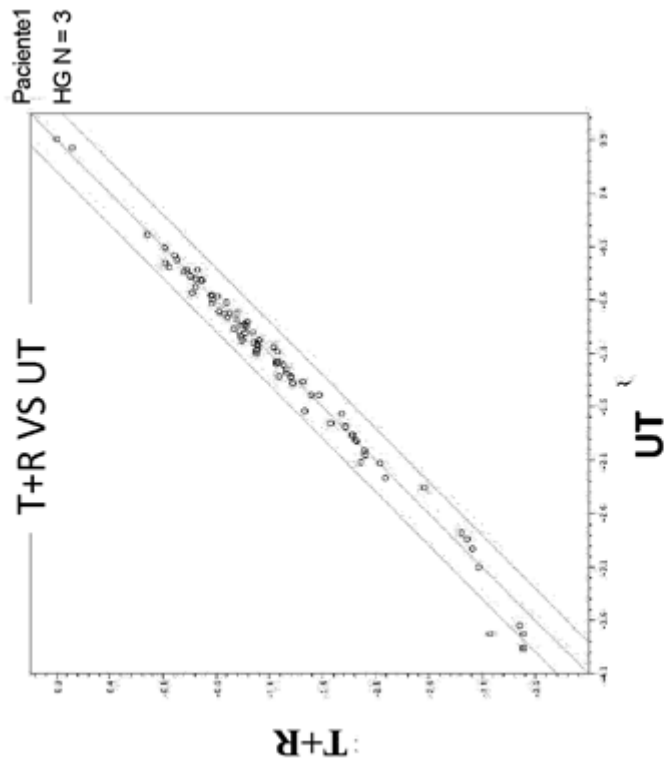


Figura 35

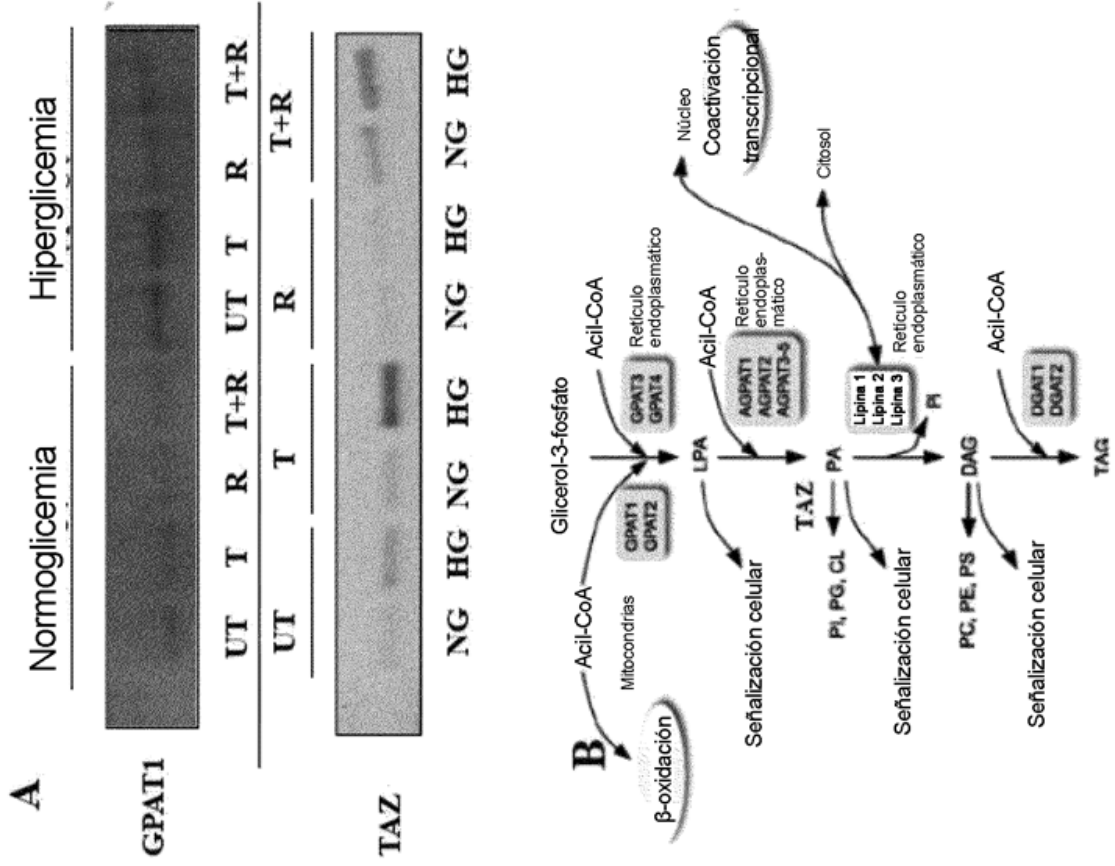


Figura 36

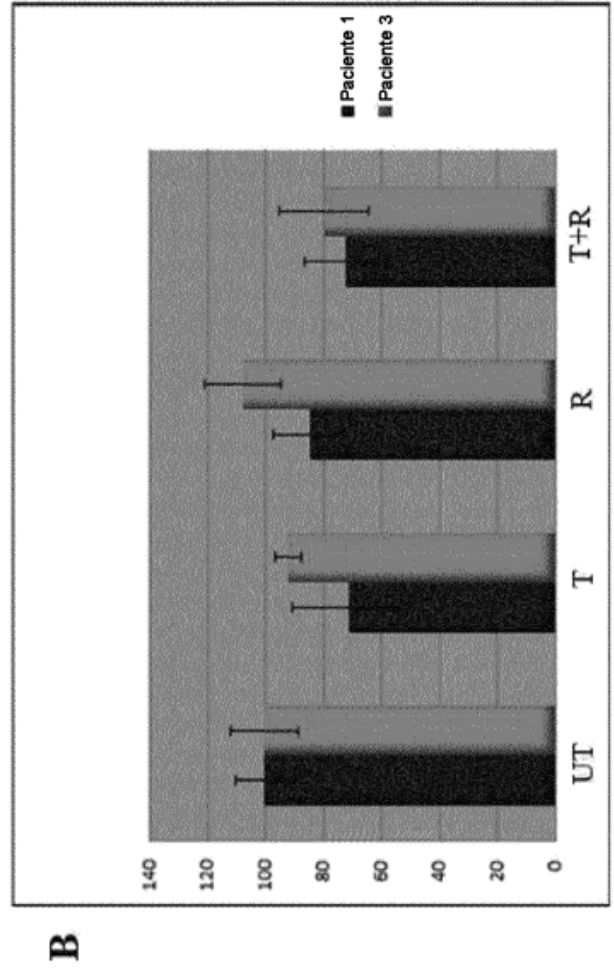
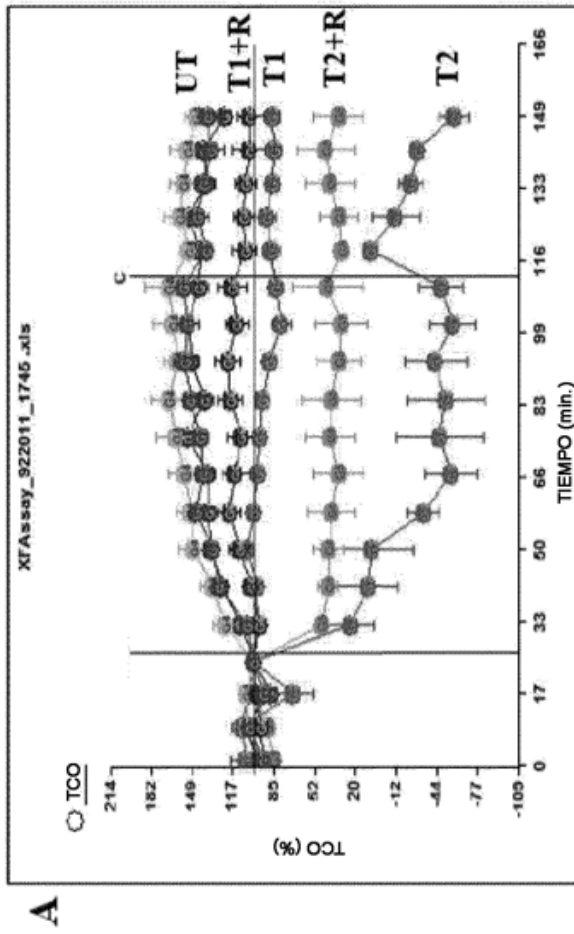


Figura 37

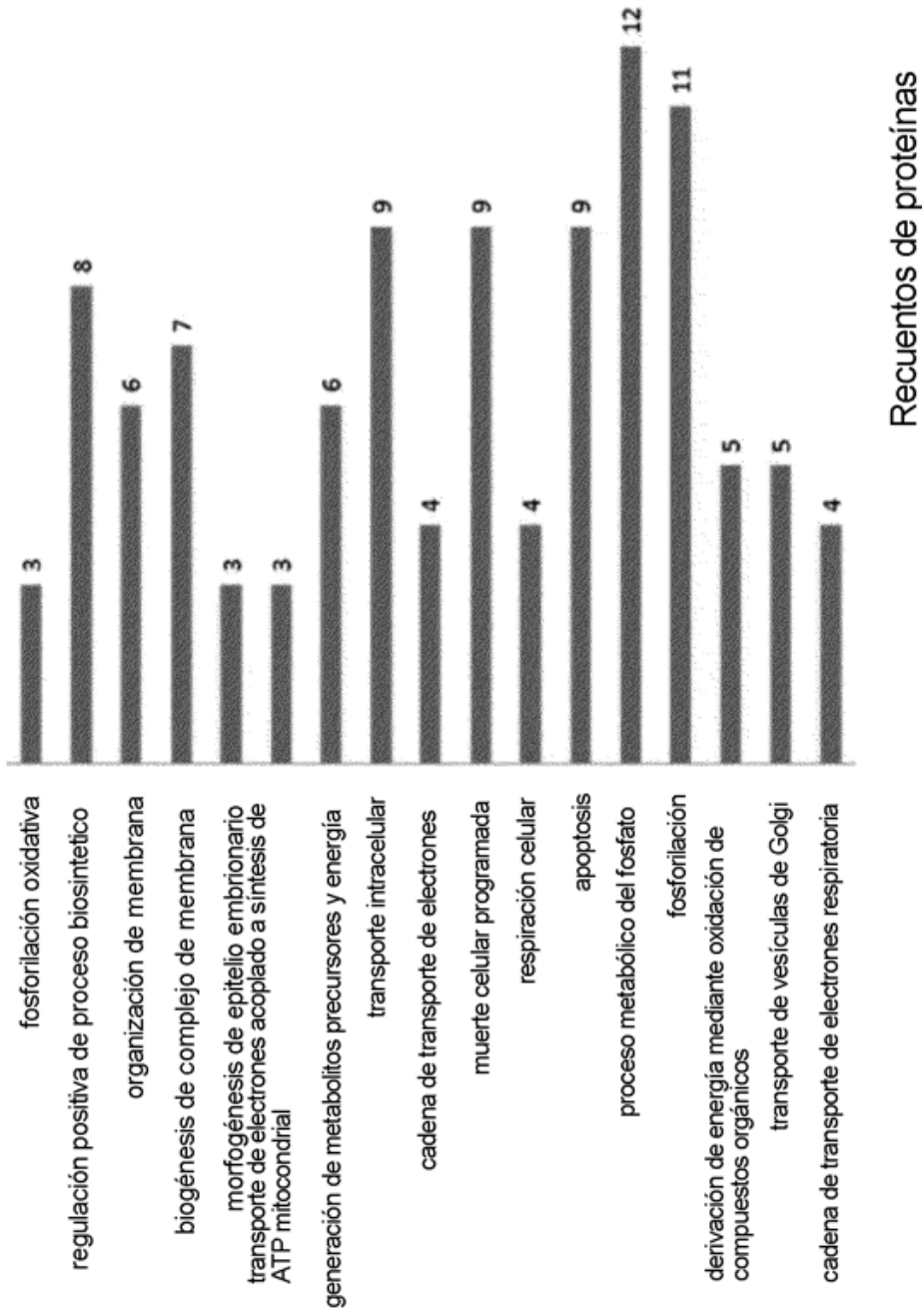


Figura 38

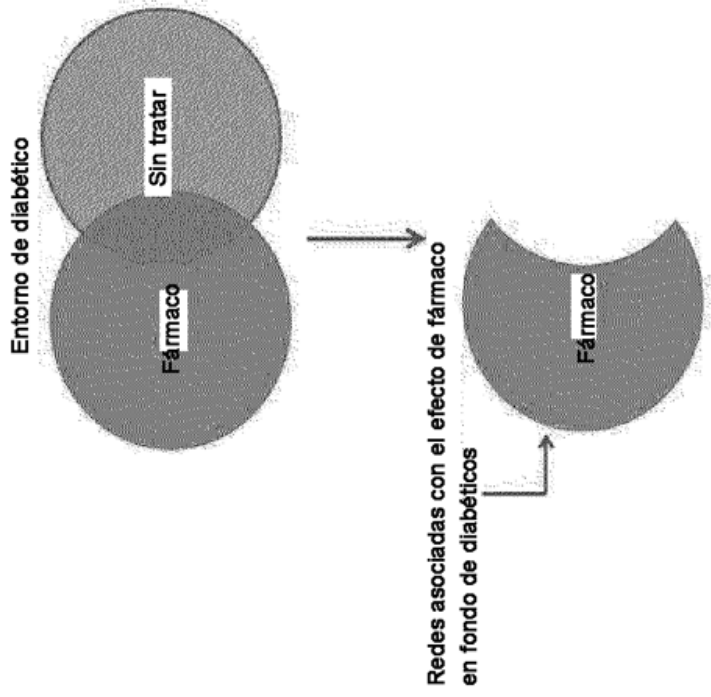


Figura 39

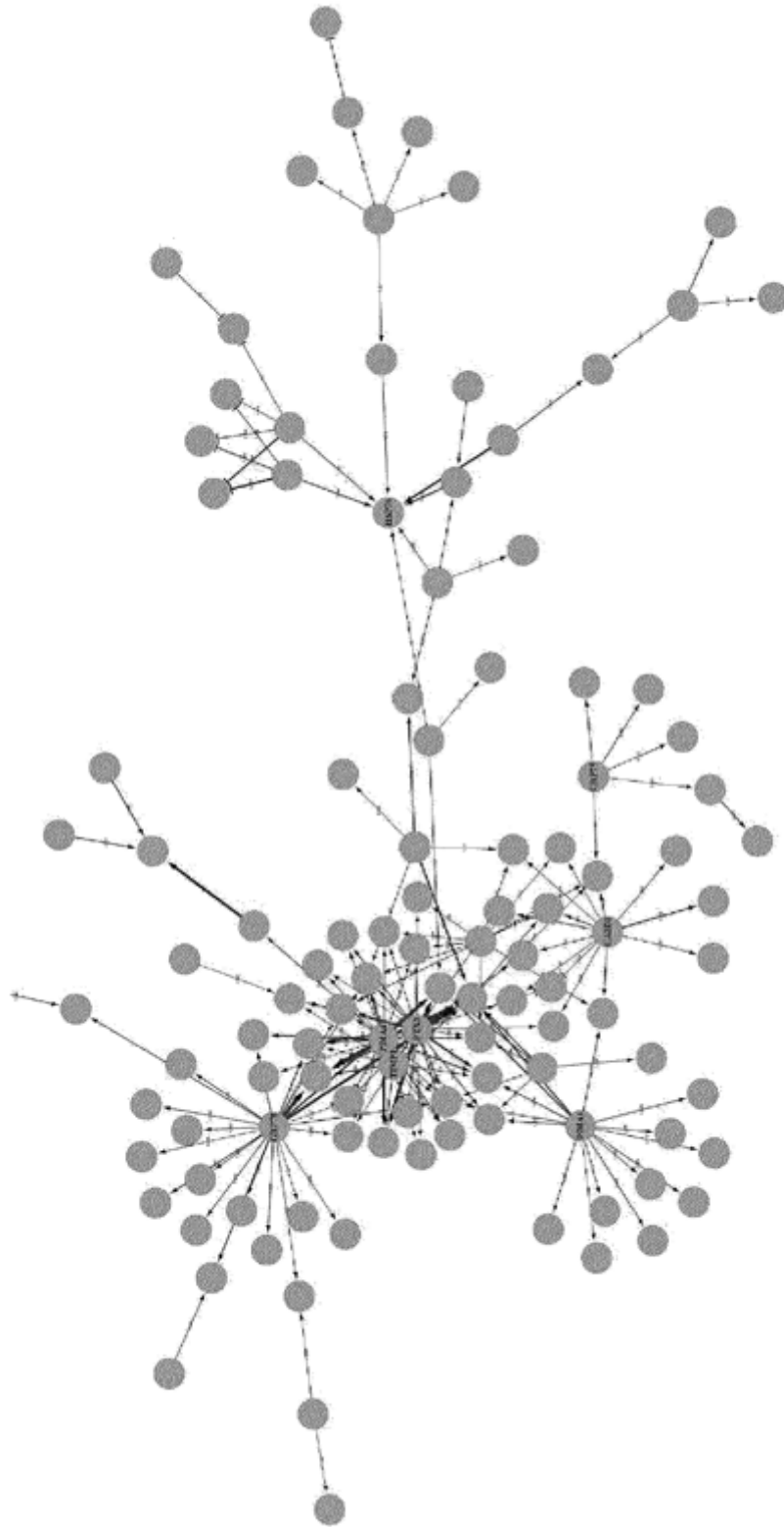


Figura 40