

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 833 556**

51 Int. Cl.:

**G01N 33/50** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.06.2016 E 19192403 (4)**

97 Fecha y número de publicación de la concesión europea: **07.10.2020 EP 3627153**

54 Título: **Un método para evaluar el estado de modificación de proteínas e identificar biomarcadores vinculados a rutas de señalización celular**

30 Prioridad:

**05.06.2015 GB 201509961**

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

**15.06.2021**

73 Titular/es:

**KINOMICA LIMITED (100.0%)  
Biohub Alderley Park, Alderley Edge,  
Macclesfield  
Cheshire SK10 4TG, GB**

72 Inventor/es:

**CUTILLAS, PEDRO RODRIGUEZ y  
WILKES, EDMUND**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 833 556 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

Un método para evaluar el estado de modificación de proteínas e identificar biomarcadores vinculados a rutas de señalización celular

5

Campo de la invención

La presente invención se refiere a un método para crear un conjunto de datos de sitios de modificación que puede usarse para preparar una base de datos de sitios de modificación de proteínas, por ejemplo sitios de fosforilación que están vinculados a rutas de señalización. Esto es útil porque proporciona información sobre qué sitios se modifican en respuesta a la modulación por fármacos particulares, tal como inhibidores. Una base de datos que contiene los datos se puede usar como entrada para otros métodos para analizar datos tales como los datos fosfoproteómicos y se puede usar para identificar marcadores de quinasas que están inhibidos por compuestos particulares o están asociados con fenotipos particulares.

15

Antecedentes de la invención

Las rutas de señalización celular forman redes complejas de reacciones bioquímicas que integran y descodifican señales extracelulares en respuestas apropiadas. La reconstrucción de estas redes y los análisis sistemáticos de sus propiedades, son importantes en el avance de nuestra comprensión molecular de la enfermedad a nivel de sistemas. La topología y la plasticidad de las redes de señalización celular desempeñan un papel importante en la fisiología fundamental y de la enfermedad. Los intentos de caracterizar dicha organización molecular se han basado en algoritmos de inferencia que obtienen información sobre las interacciones y la modificación postraduccional de proteínas (PTM) de la literatura (Kholodenko BN, *et al.* (2002) Proc. Natl. Acad. Sci. U.S.A. 99(20): 12841-12846; Mukherjee S & Speed TP (2008) Proc. Natl. Acad. Sci. U.S.A. 105(38):14313-14318; Prill RJ, *et al.* (2011) Sci. Signal 4(189):mr7). La precisión de la reconstrucción de la red utilizando tales modelos está limitada por la disponibilidad de datos (Prill RJ, *et al.*, *supra*), y por el hecho de que los eventos de señalización a menudo son específicos del tipo de célula. Como resultado, aunque pueden proporcionar datos esclarecedores, los modelos que derivan topologías de red a partir de estudios que han utilizado diferentes tipos de células y organismos dan como resultado redes compuestas o promediadas, que, de manera crítica, no siempre reflejan la estructura de la red en tipos de células específicos, en etapas específicas de desarrollo celular, o en condiciones fisiológicas definidas (Prill RJ, *et al.*, *supra*).

20

Pan *et al.* (2009) Molecular & Cellular Proteomics 8(12): 2796 - 2808 describe los efectos globales de los inhibidores de quinasas en las redes de señalización reveladas por la fosfoproteómica cuantitativa. Rix *et al.* (2007) Blood 110(12): 4055 - 4063 describe perfiles proteómicos químicos de los inhibidores de BCR-ABL imatinib, nilotinib y dasatinib que revelan nuevas dianas de quinasas y no quinasas. Vojvodic *et al.* (2013) Current Drug Discovery Technologies 10(4): 283 - 304 describe un enfoque fosfoproteómico para identificar dianas candidatas de la ruta de inhibidores de quinasas en líneas celulares primarias tipo al linfoma. Kirkpatrick *et al.* (2013) Proc. Natl. Acad. Sci. U.S.A. 110(48): 19426 - 19431 describe la caracterización fosfoproteómica de la respuesta al daño del ADN en células de melanoma después de la doble inhibición de MEK/PI3K. El documento EP2124051 describe un método para la generación rápida de perfiles de fosforilación, la detección de sitios de fosforilación *in vivo* de quinasas y fosfatasa y su uso como marcadores de diagnóstico en células, tejidos y fluidos corporales.

35

La maduración de las técnicas fosfoproteómicas basadas en la espectrometría de masas (EM) ahora permite la cuantificación simultánea de varios miles de sitios de fosforilación por experimento, y se han indicado enfoques para derivar la actividad quinasa a partir de estos conjuntos de datos de fosfoproteómica a gran escala (Linding R, *et al.* (2007) Cell 129(7): 1415-1426; Carlson SM, *et al.* (2011) Sci. Signal. 4(196); Bensimon A, *et al.* (2012) Annu. Rev. Biochem. 81:379-405; Casado P, *et al.* (2013) Sci. Signal. 6(268):rs6).

45

Uno de estos enfoques, denominado Análisis de Enriquecimiento de Sustrato de Quinasas (KSEA por sus siglas en inglés), se basa en la premisa de que, como cada sitio de fosforilación es el resultado de la actividad catalítica de una quinasa, el perfil fosfoproteómico proporciona un medio para capturar y medir las actividades de todas las quinasas expresadas en el sistema en investigación (Casado P, *et al.*, *supra*). KSEA es una técnica para inferir sistemáticamente la activación de la ruta de proteínas quinasas a partir de datos de fosfoproteómica basada en EM y es el objeto de la solicitud de patente de los presentes inventores publicada como el documento WO 2013/132075. Sin embargo, el método KSEA se basa en una recopilación de sitios de fosforilación conocidos de la literatura y de bases de datos tales como PhosphoSite (Hornbeck *et al.*, Proteomics 4, 1551 (Jun, 2004)) y PhosphoElm (Dinkel *et al.*, Nucleic Acids Res 39, D261 (enero, 2011)).

50

Existe una necesidad en la técnica de un método para analizar redes de señalización sin el conocimiento previo de sitios de fosforilación en proteínas particulares.

60

Sumario de la invención

Los presentes inventores utilizaron primero la fosfoproteómica basada en EM para definir una red de señalización de quinasas mediante la identificación sistemática de sitios de fosforilación cadena abajo de quinasas dirigidas por

65

inhibidores de quinasas de molécula pequeña de los ejes de señalización PI3K/Akt/mTOR y MEK-ERK. Estas dos vías ubicuas forman una red que regula la señalización de factores de crecimiento, antígenos e insulina, a la vez que se desregula en la mayoría de los cánceres. Después, los presentes inventores midieron la actividad y la plasticidad de diferentes rutas dentro de esta red de señalización de quinasas definida experimentalmente en células tratadas crónicamente con inhibidores de quinasas de moléculas pequeñas en estas rutas. Se descubrió que la remodelación de las redes de quinasas en células resistentes producía patrones de actividad de señalización vinculados a sus fenotipos evolucionados.

Los métodos desarrollados por los presentes inventores pueden usarse para crear un conjunto de datos de sitios de fosforilación que están vinculados a rutas de señalización. Esto es útil porque proporciona información sobre qué sitios se fosforilan o desfosforilan en respuesta a la modulación por fármacos particulares, tal como inhibidores. Después, puede usarse una base de datos que contiene los datos como entrada para otros métodos para analizar datos fosfoproteómicos, por ejemplo, el método KSEA desarrollado previamente por los presentes inventores, y puede usarse para identificar marcadores de quinasas que son inhibidos por compuestos particulares. Los métodos no se limitan al análisis de sitios de fosforilación, sino que pueden usarse para crear un conjunto de datos de sitios de modificación de proteínas que pueden detectarse usando métodos basados en EM.

Aspectos de la invención para la que se busca protección se definen en las reivindicaciones.

Según un primer aspecto, la invención proporciona un método *in vitro* para crear un conjunto de datos de sitios de modificación que se ven afectados de la misma manera por los moduladores de la misma enzima modificadora de proteínas y, por lo tanto, son indicativos de la actividad de la enzima modificadora de proteínas que está afectada por los moduladores, que comprende identificar y/o cuantificar sitios de modificación en péptidos modificados en una primera muestra que se ha tratado con un primer modulador de una enzima modificadora de proteínas y una segunda muestra que se ha tratado con un segundo modulador de la misma enzima modificadora de proteínas usando espectrometría de masas (EM); y agrupar sitios de modificación en péptidos modificados de la primera muestra que se ha tratado con el primer modulador de una enzima modificadora de proteínas y sitios de modificación en péptidos modificados de la segunda muestra que se ha tratado con el segundo modulador de la misma enzima modificadora de proteínas en un solo grupo, de acuerdo con el efecto de dichos primer y segundo moduladores de dicha enzima modificadora de proteínas sobre dichos sitios de modificación, en donde los sitios de modificación en péptidos modificados de una primera muestra y los sitios de modificación en péptidos modificados de una segunda muestra se colocan en un solo grupo basado en los sitios de modificación que se ven afectados de la misma manera por los moduladores primero y segundo de la enzima modificadora de proteínas, y en donde dichos primer y segundo moduladores de dicha enzima modificadora de proteínas son diferentes.

Dichos sitios de modificación pueden seleccionarse del grupo que consiste en sitios de fosforilación, sitios de acetilación, sitios de glicosilación, sitios de metilación y sitios de lipidación. Dicha primera muestra y dicha segunda muestra pueden ser de:

- (a) un fluido corporal, opcionalmente orina o sangre; o
- (b) un tejido, opcionalmente un tejido canceroso, opcionalmente un tumor.

Dicha enzima modificadora de proteínas puede seleccionarse del grupo que consiste en una proteína quinasa, proteína fosfatasa, proteína glicosiltransferasa, proteína acetiltransferasa, proteína metiltransferasa y proteína palmitoiltransferasa. Dicho modulador de una enzima modificadora de proteínas puede ser:

- (a) una molécula pequeña, ARNi, péptido terapéutico o anticuerpo; y/o
- (b) un inhibidor de una enzima modificadora de proteínas.

Dicho inhibidor de una enzima modificadora de proteínas puede ser un inhibidor de quinasas. Dicho inhibidor de quinasas puede ser un inhibidor de una quinasa seleccionada del grupo que consiste en Akt, CAMK2, EGFR, ERK, MEK, mTOR, p70S6K, PI3K, PKC y ROCK.

Dicho efecto de dichos primer y segundo moduladores de dicha enzima modificadora de proteínas en dicho sitio de modificación puede ser una reducción o aumento en la abundancia de dicho sitio de modificación en comparación con una muestra de control.

De acuerdo con un segundo aspecto, la invención proporciona un método para preparar una base de datos, que comprende crear un conjunto de datos de sitios de modificación de acuerdo con un método del primer aspecto y recopilar dicho conjunto de datos en una base de datos. Dicha base de datos puede comprender además información sobre las identidades de proteínas que contienen los sitios de modificación, el tipo de modificación, el tipo de muestra en la que está presente el sitio de modificación y/o el modulador que aumenta o disminuye la modificación en el sitio de modificación.

De acuerdo con un tercer aspecto, la invención proporciona un método *in vitro* para diagnosticar una enfermedad que comprende:

(a) Crear un conjunto de datos de prueba de sitios de modificación:

- i. Tratando una o más muestras de prueba de un sujeto con uno o más moduladores de una enzima modificadora de proteínas,
- ii. Identificando y/o cuantificando sitios de modificación en péptidos modificados en una o más muestras de prueba,
- iii. Agrupando sitios de modificación en péptidos modificados de una o más muestras de prueba en un solo grupo, creando así el conjunto de datos de prueba de sitios de modificación, y
- iv. opcionalmente recopilando el conjunto de datos de prueba en una base de datos de prueba,

(b) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el método del primer aspecto de la invención o comparar la base de datos de prueba con una base de datos preparada de acuerdo con el segundo aspecto de la invención,

(c) Encontrar una diferencia significativa entre el conjunto de datos de prueba o la base de datos de prueba y el conjunto de datos de control o la base de datos de control,

(d) Diagnosticar que el sujeto tiene la enfermedad si se encuentra una diferencia significativa entre el conjunto de datos o la base de datos en la etapa (c).

La enfermedad puede ser cáncer.

De acuerdo con un cuarto aspecto, la invención proporciona un método para identificar uno o más biomarcadores para determinar la actividad de una enzima modificadora de proteínas que comprende crear un conjunto de datos de sitios de modificación de acuerdo con el método del primer aspecto de la invención o preparar una base de datos de acuerdo con el segundo aspecto de la invención, en donde los sitios de modificación en el único grupo se identifican como el uno o más biomarcadores para la actividad de una enzima modificadora de proteínas.

De acuerdo con un quinto aspecto, la invención proporciona un método *in vitro* para determinar que una sustancia de prueba es un modulador de una enzima modificadora de proteínas que comprende tratar una muestra de prueba con la sustancia de prueba, identificar y/o cuantificar sitios de modificación en péptidos modificados en la muestra de prueba, crear un conjunto de datos de prueba de sitios de modificación a partir de los sitios de modificación identificados y/o cuantificados en la muestra de prueba, que comprende además:

(a) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el método del primer aspecto de la invención y/o

(b) Recopilar dicho conjunto de datos de prueba en una base de datos de prueba y comparar la base de datos de prueba con una base de datos preparada de acuerdo con el segundo aspecto de la invención.

Descripción detallada de la invención

En el presente documento se describe un método para crear un conjunto de datos de sitios de modificación mediante el análisis de sitios de modificación en péptidos modificados a partir de una primera y una segunda muestra. La mayoría de las proteínas y péptidos se modifican de alguna manera mediante la adición o eliminación de grupos funcionales y dichas modificaciones se efectúan mediante enzimas modificadoras de proteínas. Las modificaciones de proteínas y péptidos que pueden detectarse por espectrometría de masas (EM) incluyen fosforilación, glicosilación, acetilación, metilación y lipidación. Dichas modificaciones son ejemplos de modificaciones postraduccionales y tienen diversas funciones biológicas en la célula. Por lo tanto, los sitios de modificación pueden, por lo tanto, ser sitios de modificaciones postraduccionales. Por ejemplo, los sitios de modificación pueden ser sitios de fosforilación, glicosilación, acetilación, metilación y lipidación. Los sitios de modificación son generalmente sitios de modificación de proteínas y/o péptidos. Un sitio de modificación puede ser uno o más restos de aminoácidos de un péptido o proteína a los que se añade un grupo funcional tal como un grupo fosfato al péptido o proteína. Los grupos funcionales alternativos incluyen carbohidratos, grupos acetilo, grupos metilo y lípidos. Un "péptido modificado" se define en el presente documento, como un péptido que se ha modificado mediante la adición o eliminación de un grupo funcional. Una "enzima modificadora de proteínas" se define, en el presente documento, como una enzima que cataliza una reacción que implica la adición o eliminación de un grupo funcional a una proteína o péptido.

Un "péptido" como se define en el presente documento es una secuencia de aminoácidos corta e incluye oligopéptidos y polipéptidos. Generalmente, tales péptidos tienen entre aproximadamente 5 y 30 aminoácidos de longitud, por ejemplo de 6 o 7 a 25, 26 o 27 aminoácidos, de 8, 9 o 10 a 20 aminoácidos, de 11 o 12 a 18 aminoácidos o de 14 a 16 aminoácidos, por ejemplo 15 aminoácidos. Sin embargo, también se pueden usar péptidos más cortos y más largos, tales como entre aproximadamente 2 y aproximadamente 50, por ejemplo, de aproximadamente 3 a aproximadamente 35 o 40 o de aproximadamente 4 a aproximadamente 45 aminoácidos. Generalmente, el péptido es adecuado para el análisis espectrométrico de masas, es decir, la longitud del péptido es tal que el péptido es adecuado para el análisis espectrométrico de masas. La longitud del péptido que puede analizarse está limitada por la capacidad del espectrómetro de masas para secuenciar péptidos tan largos. En determinados casos, se pueden analizar polipéptidos de hasta 300 aminoácidos, por ejemplo de 50 a 250 aminoácidos, de 100 a 200 aminoácidos o de 150 a 175 aminoácidos.

El método se puede aplicar al análisis de cualquier modificación postraduccional que se pueda detectar utilizando métodos basados en EM. Dichas modificaciones incluyen las que están mediadas por enzimas que incluyen proteínas quinasas, proteínas fosfatasa, proteínas glicosiltransferasas, proteínas acetiltransferasas, proteínas metiltransferasas y proteínas palmitoiltransferasas. La actividad de estas enzimas produce fosforilación, desfosforilación, acetilación, glicosilación, metilación y lipidación de sustratos de proteínas o péptidos respectivamente. Todas estas modificaciones de proteínas/péptidos pueden detectarse por espectrometría de masas.

Por consiguiente, los péptidos modificados para su uso en el método son generalmente péptidos fosforilados, péptidos desfosforilados, péptidos acetilados, péptidos glicosilados, péptidos metilados y/o péptidos lipidados. Los péptidos modificados contienen uno o más aminoácidos con una modificación postraduccional, que es específica de la enzima modificadora de proteínas que provoca la modificación. Dichos aminoácidos modificados se denominan en el presente documento "sitios de modificación".

Los péptidos modificados para su uso en el método pueden ser péptidos fosforilados. Los péptidos fosforilados contienen uno o más aminoácidos que están fosforilados (es decir, se ha agregado un grupo fosfato (PO<sub>4</sub>) a ese aminoácido). Dichos aminoácidos fosforilados se denominan en el presente documento "sitios de fosforilación". En relación con esto, el término "fosfoproteína" se usa en el presente documento para referirse a una proteína fosforilada y el término "fosfopéptido" se usa en el presente documento para referirse a un péptido fosforilado.

Las proteínas quinasas humanas se pueden dividir en varios grupos, incluidas las quinasas AGC, por ejemplo, la proteína quinasa A (PKA), la proteína quinasa B (PKB) (también conocida como Akt), la proteína quinasa C (PKC) y la proteína quinasa G (PKG); tirosina quinasas tales como receptores tirosina quinasas (por ejemplo, el receptor del factor de crecimiento epidérmico (EGFR)); tirosina quinasas tipo quinasas; proteínas quinasas dependientes de calcio/calmodulina (por ejemplo CAMK2); el grupo caseína quinasa 1; el grupo CMGC, por ejemplo, CDK, MAPK (tales como quinasas relacionadas con señales extracelulares (ERK), quinasas de proteínas quinasas activadas por mitógeno (por ejemplo MEK)), quinasas GSK3 y CLK; y STE, los homólogos de quinasas de levadura Sterile 7, Sterile 11 y Sterile 20. Otras quinasas incluyen mTOR (diana de rapamicina en mamíferos), quinasa p70S6 y proteína quinasa asociada a rho (ROCK), todas las cuales son serina/treonina proteína quinasas y fosfatidilinositol-4,5-bisfosfato 3-quinasa (PI3K).

Las proteínas fosfatasa incluyen la proteína fosfatasa 2 (PP2 o PP2A).

El método implica agrupar sitios de modificación en péptidos modificados de una primera muestra y sitios de modificación en péptidos modificados de una segunda muestra en un solo grupo. Las muestras primera y segunda utilizadas en los métodos pueden ser cualquier muestra que contenga péptidos. La muestra generalmente es una muestra biológica y, por lo tanto, puede ser cualquier tipo de muestra obtenida de una fuente biológica, por ejemplo, una muestra obtenida de un ser humano, animal, vegetal o bacteria. Las muestras utilizadas pueden obtenerse de fuentes humanas y no humanas.

Las muestras utilizadas en los métodos pueden ser de cualquier especie de interés. Generalmente, las muestras son de un ser humano o animal. El animal generalmente es un mamífero, por ejemplo un roedor tal como un ratón, rata o cobaya, o un ungulado tal como una vaca, oveja o cabra. El animal es alternativamente un pájaro, tal como un pollo, un pez, tal como un pez cebra, un nematodo, tal como el gusano *Caenorhabditis elegans*, o un insecto, tal como la mosca de la fruta *Drosophila melanogaster*. Las muestras utilizadas en los métodos también pueden ser de otras formas de vida, como bacterias y levaduras. Las muestras utilizadas en los métodos generalmente son muestras de una especie de bacteria experimentalmente importante como *Escherichia coli*, *Salmonella enterica*, *Streptococcus pneumoniae* o *Staphylococcus aureus*, o de levadura tal como la levadura de panadería *Saccharomyces cerevisiae* o la levadura de fisión *Schizosaccharomyces pombe*. Las muestras utilizadas en los métodos pueden ser alternativamente de una planta u hongo o un virus.

Generalmente, la muestra biológica deriva de un ser humano y puede ser, por ejemplo, una muestra de un fluido corporal tal como orina o sangre u otro tejido. Generalmente, la muestra biológica es una célula de una línea celular o un tejido, generalmente un tejido primario. Por ejemplo, la muestra puede ser un tejido de un ser humano o animal. El ser humano o animal puede estar sano o enfermo. El tejido puede ser tejido canceroso. Por ejemplo, la muestra puede ser de un tumor. Alternativamente, la muestra puede ser una línea celular derivada de células humanas o animales sanas o enfermas.

La primera y la segunda muestra utilizadas en el método generalmente son o incluyen células de líneas celulares, por ejemplo una línea celular de cáncer tal como una línea celular de cáncer de mama, por ejemplo una línea celular MCF7. Muchas líneas celulares de cáncer se conocen en la técnica y se enumeran en línea, por ejemplo, en la Enciclopedia de Líneas Celulares de Cáncer de Broad-Novartis (CCLE) en <http://www.broadinstitute.org/ccle/home>

La primera y la segunda muestra son generalmente de la misma línea celular pero pueden ser diferentes. La primera y segunda muestras pueden proceder de la misma fuente. Por ejemplo, la primera y la segunda muestra pueden ser ambas de un solo individuo. La primera y la segunda muestra pueden ser del mismo tejido. La primera y la segunda muestra pueden ser del mismo fluido corporal.

Una muestra puede denominarse en el presente documento, una "muestra de prueba" para distinguir la muestra de otra muestra utilizada en un método. Por ejemplo, una muestra se puede denominar muestra de prueba para describir un método que implique una comparación con una muestra de control.

5 El método es un método *in vitro* y, por lo tanto, no comprende la etapa de obtener una muestra de un organismo tal como un animal.

Los péptidos modificados pueden modificarse en la muestra. Los péptidos modificados pueden modificarse *in vivo*. Los péptidos modificados pueden ser péptidos modificados endógenos.

10 En la presente invención, la primera y la segunda muestras se han tratado con un modulador de una enzima modificadora de proteínas antes de llevar a cabo el método de la invención. En la práctica, esto significa que la primera y la segunda muestras en sí mismas o el organismo del que se obtienen la primera y la segunda muestras se tratan con un modulador de una enzima modificadora de proteínas antes de llevar a cabo el método de la invención.

15 El método también puede comprender la etapa de obtener péptidos modificados a partir de la primera y/o segunda muestra. Por ejemplo, cuando la primera y/o segunda muestra es o incluye una célula (por ejemplo, a partir de una línea celular), los péptidos modificados pueden obtenerse a partir de la célula lisando (apertura por fraccionamiento) la célula, extrayendo proteínas de las células lisadas y escindiendo (descomponiendo o digiriendo) las proteínas en péptidos. Las células se pueden lisar usando cualquier medio adecuado conocido en la técnica, por ejemplo, usando métodos físicos tales como lisis mecánica (por ejemplo, usando un mezclador Waring), homogeneización de líquidos, sonicación o lisis manual (por ejemplo, usando una mano de mortero y mortero) métodos basados en detergentes tal como CHAPS o Triton-X. Generalmente, las células se lisan usando un tampón desnaturizante tal como un tampón basado en urea. La escisión o digestión de proteínas puede llevarse a cabo usando cualquier agente adecuado conocido en la técnica y generalmente se lleva a cabo usando una proteasa. Se puede usar cualquier proteasa adecuada. La proteasa generalmente es tripsina, quimiotripsina, Arg-C, pepsina, V8, Lys-C, Asp-C y/o AspN. Alternativamente, las proteínas se pueden escindir químicamente, por ejemplo utilizando hidroxilamina, ácido fórmico, bromuro de cianógeno, BNPS-escatol, ácido 2-nitro-5-tiocianobenzoico (NTCB) o cualquier otro agente adecuado.

20 La enzima modificadora de proteínas generalmente se selecciona del grupo que consiste en una proteína quinasa, proteína fosfatasa, proteína glicosiltransferasa, proteína acetiltransferasa, proteína metiltransferasa y proteína palmitoiltransferasa. En una realización, la enzima modificadora de proteínas es una proteína quinasa.

35 Los moduladores de una enzima modificadora de proteínas utilizados para tratar la primera y segunda muestra antes de llevar a cabo el método son moduladores de la misma enzima modificadora de proteínas, pero los moduladores utilizados para tratar la primera y segunda muestra son diferentes, generalmente estructuralmente diferentes.

40 El modulador de una enzima modificadora de proteínas es generalmente un químico o fármaco exógeno, tal como una molécula pequeña (tal como un inhibidor de molécula pequeña), ARNi, péptido terapéutico o anticuerpo.

45 El modulador de una enzima modificadora de proteínas puede modular la actividad de una enzima modificadora de proteínas de cualquier manera. El modulador de una enzima modificadora de proteínas puede ser un inhibidor de una enzima modificadora de proteínas. El inhibidor de una enzima modificadora de proteínas puede ser un inhibidor de quinasas. El modulador de una enzima modificadora de proteínas puede ser un activador de una enzima modificadora de proteínas. El inhibidor de una enzima modificadora de proteínas puede ser un activador de quinasas.

50 Los inhibidores de quinasas típicos incluyen inhibidores de Akt, tal como el inhibidor de Akt VIII y MK-2206, inhibidores de CAMK2, tales como KN-93 y KN-62, inhibidores de EGFR, tales como PD-168393 y PD-153035, inhibidores de ERK, tales como inhibidor de ERK e inhibidor de ERK II, inhibidores de MEK, tales como GSK-1120212 y U0126, inhibidores de mTOR, tales como KU-0063794 y Torin-1, inhibidores de p70S6K, tales como PF-4708671 y DG2, inhibidores de PI3K, tales como GDC-0941 y PI-103, inhibidores de PKC, tales como Gö-6976 y BIM-1 e inhibidores de ROCK, tales como H-1152 e Y-27632.

55 Otros inhibidores de PI3K incluyen PP2 y wortmannin. Al menos 80 inhibidores de quinasas se encuentran en diferentes etapas de desarrollo clínico (Zhang, J.; *et al.* Nat Rev Cancer 2009, 9, (1), 28-39).

El inhibidor de una enzima modificadora de proteínas puede ser un inhibidor de fosfatasas. Los inhibidores de fosfatasas incluyen los inhibidores de PP2A, tales como el ácido ocadaico y el ácido cantárico.

60 El método comprende agrupar sitios de modificación en péptidos modificados de una primera muestra que se ha tratado con el primer modulador de una enzima modificadora de proteínas y sitios de modificación en péptidos modificados de una segunda muestra que se ha tratado con un segundo modulador de la misma enzima modificadora de proteínas en un solo grupo. Por "agrupar" se entiende que los sitios de modificación tales como los sitios de fosforilación se colocan en un grupo o conjunto. En el método, los sitios de modificación en péptidos modificados de una primera muestra y los sitios de modificación en péptidos modificados de una segunda muestra se colocan en un solo grupo basándose en el efecto del primer y segundo modulador de la enzima modificadora de proteínas sobre los

sitios de modificación. En otras palabras, los sitios de modificación en péptidos modificados de una primera muestra y los sitios de modificación en péptidos modificados de una segunda muestra se colocan en un solo grupo basándose en los sitios de modificación que se ven afectados de la misma manera por el primer y segundo modulador de la enzima modificadora de proteínas.

Por ejemplo, el efecto del primer y segundo modulador de la enzima modificadora de proteínas sobre los sitios de modificación puede ser una reducción o aumento en la abundancia (frecuencia) de un sitio de modificación particular respecto a una muestra de control. En este contexto, los sitios de modificación en péptidos modificados de una primera muestra y los sitios de modificación en péptidos modificados de una segunda muestra se colocan en un solo grupo basándose en una reducción o aumento similar en la abundancia de un sitio de modificación particular en respuesta al primer y segundo modulador de la enzima modificadora de proteínas, en comparación con una muestra de control. La reducción o aumento de la abundancia es generalmente una reducción o aumento estadísticamente significativo y puede calcularse usando cualquier método estadístico adecuado, por ejemplo, como se describe en el Ejemplo en el presente documento.

El resultado final del método es un conjunto de datos de sitios de modificación que se ven afectados de la misma manera por los moduladores de la misma enzima modificadora de proteínas y, por lo tanto, son indicativos de la actividad de la enzima modificadora de proteínas que está afectada por el modulador. La selección de sitios de modificación inhibidos por moduladores diferentes (y generalmente estructuralmente distintos) que se dirigen a la misma enzima modificadora de proteínas da como resultado conjuntos de datos enriquecidos en sitios de modificación específicos para la enzima modificadora de proteínas prevista. Si un sitio de modificación particular en un péptido modificado se reduce en abundancia por al menos dos moduladores diferentes de la misma enzima modificadora de proteínas, entonces esto representa una lectura de la enzima modificadora de proteínas que en realidad se ve afectada por el modulador. Dichos sitios de modificación se denominan en el presente documento "marcadores de actividad compuesto-diana" (CTAM). Por consiguiente, dichos sitios de modificación pueden considerarse biomarcadores para la actividad de una enzima modificadora de proteínas.

El método puede implicar agrupar sitios de modificación en péptidos modificados de más de dos muestras, por ejemplo 3, 4, 5, 6, 7, 8, 9, 10 o incluso más muestras, cada una de las cuales se ha tratado con un modulador diferente de la misma enzima modificadora de proteínas. En este contexto, el método implica agrupar sitios de modificación en péptidos modificados de todas las muestras en un solo grupo de acuerdo con el efecto de cada uno de los moduladores de la misma enzima modificadora de proteínas en los péptidos modificados, como se describe en el presente documento. Por ejemplo, cuando el método implica 3 muestras, el método implica agrupar sitios de modificación en péptidos modificados de una primera muestra que se ha tratado con un primer modulador de una enzima modificadora de proteínas, sitios de modificación en péptidos modificados de una segunda muestra que se ha tratado con una segunda modulador de la misma enzima modificadora de proteínas y sitios de modificación en péptidos modificados de una tercera muestra que se ha tratado con un tercer modulador de la misma enzima modificadora de proteínas en un solo grupo, de acuerdo con el efecto de dichos primero, segundo y tercer moduladores de dicho enzima modificadora de proteínas en dichos sitios de modificación, en donde dichos primer, segundo y tercer moduladores de dicha enzima modificadora de proteínas son diferentes. Por ejemplo, el efecto del primer, segundo y tercer modulador de la enzima modificadora de proteínas en los sitios de modificación puede ser una reducción o aumento en la abundancia (número) de un sitio de modificación particular respecto a una muestra de control. En este contexto, los sitios de modificación en péptidos modificados de una primera muestra, los sitios de modificación en péptidos modificados de una segunda muestra y los sitios de modificación en péptidos modificados de una tercera muestra se colocan en un solo grupo basándose en una reducción o aumento similar en la abundancia de un sitio de modificación particular en respuesta al primer, segundo y tercer modulador de la enzima modificadora de proteínas, en comparación con una muestra de control. Cuando se utilizan más de dos muestras, puede haber una única muestra de control con la que se comparan cada una de las muestras de prueba.

El resultado del método es un conjunto de datos de sitios de modificación (por ejemplo, sitios de fosforilación) que se ven afectados por los moduladores de las enzimas modificadoras de proteínas, tal como los inhibidores de las proteínas quinasas. El conocimiento de dichos sitios de modificación permite la vinculación de los inhibidores a las actividades de las quinasas y al fenotipo de las células. El conocimiento de dichos sitios de modificación también proporciona información sobre los circuitos de las rutas de señalización y, cuando se monitorizan en otros experimentos, dichos sitios de modificación también pueden revelar la plasticidad de las rutas de señalización; es decir, cómo cambian en determinadas condiciones experimentales o durante la progresión de la enfermedad. El método también se puede utilizar para identificar nuevos marcadores o biomarcadores y, por lo tanto, proporcionar información sobre la actividad de las quinasas y otras enzimas modificadoras de proteínas dirigidas por inhibidores específicos.

El método, el conjunto de datos creado de acuerdo con el método y/o la base de datos preparada de acuerdo con el método también se puede usar en un método para diagnosticar una enfermedad. La enfermedad puede ser cáncer. La primera muestra y la segunda muestra pueden ser del sujeto. La primera muestra y la segunda muestra pueden ser de un tumor del sujeto. La primera muestra y la segunda muestra pueden ser de diferentes etapas de progresión de la enfermedad. El método de diagnóstico puede comprender administrar uno o más moduladores de una enzima modificadora de proteínas a la primera muestra y a la segunda muestra *in vitro*. El método puede permitir la

identificación de los mecanismos de resistencia terapéutica relevantes para cada paciente individual. El método de diagnóstico puede comprender además evaluar las respuestas celulares al uno o más moduladores de la enzima modificadora de proteínas. Las respuestas celulares pueden ser, por ejemplo, proliferación o viabilidad celular. Las respuestas celulares pueden estar correlacionadas con el conjunto de datos y/o la base de datos. El método de diagnóstico puede comprender la comparación del conjunto de datos o de la base de datos obtenidos de las muestras del sujeto con una o más muestras de control. Generalmente, las muestras de control serán de tejido sano o de un sujeto sano. Las muestras de control pueden ser de tejido sano o enfermo o de un sujeto sano o enfermo. La comparación con las muestras de control puede indicar que una o más enzimas modificadoras de proteínas están desreguladas en el sujeto respecto al control. El método de diagnóstico puede comprender además la identificación de dianas terapéuticas relevantes para el sujeto. El método de diagnóstico puede comprender además determinar un modulador eficaz de una enzima modificadora de proteínas.

El método, el conjunto de datos creado de acuerdo con el método y/o la base de datos preparada de acuerdo con el método también se pueden usar en un método para determinar que una sustancia de prueba es un modulador de una enzima modificadora de proteínas. Por consiguiente, el método, el conjunto de datos creado de acuerdo con el método y/o la base de datos preparada de acuerdo con el método también se puede usar en un método de detección. El método de detección puede ser un método de detección de sustancias de prueba para determinar si la sustancia de prueba es un modulador de una enzima modificadora de proteínas. La sustancia de prueba puede ser una sustancia química o fármaco exógenos, tal como una molécula pequeña (tal como un inhibidor de molécula pequeña), ARNi, un péptido terapéutico o un anticuerpo. La sustancia de prueba puede ser una sustancia nueva o una sustancia conocida. La sustancia de prueba puede ser una composición.

Por lo tanto, determinar que una sustancia de prueba es un modulador de una enzima modificadora de proteínas puede implicar determinar que la sustancia de prueba es un inhibidor de una enzima modificadora de proteínas. Alternativamente, la sustancia de prueba puede ser un activador de la enzima modificadora de proteínas. Por ejemplo, la sustancia de prueba puede identificarse como un inhibidor de quinasas o un activador de quinasas.

El método se basa en el análisis de sitios de modificación en péptidos modificados identificados usando técnicas basadas en EM. Por consiguiente, los sitios de modificación en péptidos modificados de una primera muestra y de una segunda muestra que se agrupan en la primera etapa del método se identifican y/o cuantifican generalmente, usando técnicas basadas en EM. Por lo tanto, el método incluye una etapa de identificación y/o cuantificación de sitios de modificación en péptidos modificados en una primera muestra y/o una segunda muestra usando espectrometría de masas (EM), antes de la etapa de agrupar los sitios de modificación en péptidos modificados de una primera muestra y de una segunda muestra en un solo grupo, como se describe en el presente documento.

La identificación y cuantificación de sitios de modificación en péptidos modificados se puede llevar a cabo utilizando cualquier método adecuado. Generalmente, la cuantificación puede llevarse a cabo mediante cualquier método que implique espectrometría de masas (EM), tal como cromatografía líquida-espectrometría de masas (CL-EM). La CL-EM o CL-EM/EM es generalmente EM libre sin marcadores, pero también se pueden usar como base para el análisis las técnicas que usan el marcado de isótopos como base para la cuantificación.

En los métodos, la cuantificación de una modificación de proteínas tal como la fosforilación, se lleva a cabo generalmente utilizando la técnica TIQUAS (cuantificación dirigida y en profundidad de la señalización), tal como se describe en el documento WO 2010/119261 (Solicitud de Patente Internacional n.º PCT/GB2010/000770), así como en Casado P, *et al.* (2013) *Genome Biol.* 14(4):R37 y Montoya A, *et al.* (2011) *Methods* 54(4):370-378. Esta técnica permite la cuantificación sensible, rápida y completa de péptidos modificados. El método puede, en un ensayo simple, medir simultáneamente las cantidades de miles de sitios de fosforilación en proteínas. Como se expone en el documento WO 2010/119261, la técnica TIQUAS también se puede utilizar para cuantificar péptidos modificados distintos de los péptidos fosforilados. De hecho, la técnica TIQUAS puede usarse para cuantificar péptidos que contienen cualquier modificación que pueda detectarse por espectrometría de masas.

En este contexto, la etapa de cuantificar sitios de modificación en péptidos modificados usando espectrometría de masas (EM) antes de la etapa de agrupación se lleva a cabo usando un método que comprende las siguientes etapas:

- (a) obtener péptidos de una muestra;
- (b) añadir péptidos modificados de referencia a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos y péptidos modificados de referencia;
- (c) llevar a cabo una espectrometría de masas (EM) en dicha mezcla de péptidos y péptidos modificados de referencia para obtener datos relacionados con los péptidos en la muestra; y
- (d) comparar los datos relacionados con los péptidos en la muestra con los datos en una base de datos de péptidos modificados utilizando un programa informático;

en donde la base de datos de péptidos modificados se recopila mediante un método que comprende:

- i obtener péptidos de una muestra;
- ii enriquecer péptidos modificados a partir de los péptidos obtenidos en la etapa i;

- iii realizar cromatografía líquida-espectrometría de masas en tándem (CL-EM/EM) en los péptidos modificados enriquecidos obtenidos en la etapa ii;
- iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y
- v recopilar datos relacionados con los péptidos modificados identificados en la etapa iv en una base de datos.

Cuando la enzima modificadora de proteínas es una proteína quinasa y la modificación es fosforilación, la etapa de cuantificar los sitios de modificación en péptidos modificados usando espectrometría de masas (EM) antes de la etapa de agrupación puede llevarse a cabo usando un método que comprende las siguientes etapas:

- (a) obtener péptidos fosforilados de una muestra;
- (b) añadir péptidos fosforilados de referencia a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos fosforilados y péptidos fosforilados de referencia;
- (c) llevar a cabo una espectrometría de masas (EM) en dicha mezcla de péptidos fosforilados y péptidos fosforilados de referencia para obtener datos relacionados con los péptidos fosforilados en la muestra; y
- (d) comparar los datos relacionados con los péptidos fosforilados en la muestra con los datos en una base de datos de péptidos fosforilados utilizando un programa informático;

en donde la base de datos de péptidos fosforilados se recopila mediante un método que comprende:

- i obtener péptidos de una muestra;
- ii enriquecer péptidos fosforilados a partir de los péptidos obtenidos en la etapa i;
- iii realizar cromatografía líquida-espectrometría de masas en tándem (CL-EM/EM) en los péptidos fosforilados enriquecidos obtenidos en la etapa ii;
- iv comparar los péptidos fosforilados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos fosforilados; y
- v recopilar datos relacionados con los péptidos fosforilados identificados en la etapa iv en una base de datos.

En relación con este método, la palabra "péptido" se usa indistintamente con la palabra "polipéptido".

La etapa (a) de este método implica obtener péptidos de una muestra. Los péptidos se pueden obtener de la muestra usando cualquier método adecuado conocido en la técnica. La etapa (a) puede comprender:

- (1) lisar las células en la muestra;
- (2) extraer las proteínas de las células lisadas obtenidas en la etapa (1); y
- (3) escindir dichas proteínas en péptidos.

En la etapa (1), las células de la muestra se lisan o se abren por fraccionamiento. Las células se pueden lisar usando cualquier medio adecuado conocido en la técnica, por ejemplo, usando métodos físicos tales como lisis mecánica (por ejemplo, usando un mezclador Waring), homogeneización de líquidos, sonicación o lisis manual (por ejemplo, usando una mano de mortero y mortero) métodos basados en detergentes tal como CHAPS o Triton-X. Generalmente, las células se lisan usando un tampón desnaturante tal como un tampón basado en urea.

En la etapa (2), las proteínas se extraen de las células lisadas obtenidas en la etapa (1). En otras palabras, las proteínas se separan de los otros componentes de las células lisadas.

En la etapa (3), las proteínas de las células lisadas se escinden en péptidos. En otras palabras, las proteínas se descomponen en péptidos más cortos. La descomposición de proteínas también se conoce comúnmente como digestión. La escisión de proteínas se puede llevar a cabo utilizando cualquier agente adecuado conocido en la técnica.

La escisión o digestión de proteínas se lleva a cabo generalmente usando una proteasa. Se puede usar cualquier proteasa adecuada. La proteasa generalmente es tripsina, quimiotripsina, Arg-C, pepsina, V8, Lys-C, Asp-C y/o AspN. Alternativamente, las proteínas se pueden escindir químicamente, por ejemplo utilizando hidroxilamina, ácido fórmico, bromuro de cianógeno, BNPS-escatol, ácido 2-nitro-5-tiocianobenzoico (NTCB) o cualquier otro agente adecuado.

En la etapa (b) de este método, los péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) se añaden a los péptidos obtenidos en la etapa (a) para producir una mezcla de péptidos y péptidos modificados de referencia (generalmente péptidos fosforilados de referencia). La etapa (b) da como resultado una mezcla de péptidos (incluidos los modificados, generalmente los fosforilados) por muestra. Los péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) también se denominan en el presente documento "patrones internos" (IS (*internal standards*) por sus siglas en inglés). Generalmente, se añaden de 5 a 10, por ejemplo de 6 a 9 o de 7 a 8, péptidos modificados de referencia (generalmente péptidos fosforilados de referencia).

Los péptidos modificados de referencia son generalmente péptidos fosforilados de referencia y derivan generalmente de una proteína de referencia de naturaleza y concentración definidas, a menudo denominada proteína de patrón interno (IS). Los IS pueden ser proteínas disponibles comercialmente, por ejemplo, caseína. Alternativamente, los IS

- 5 se sintetizan específicamente. En este contexto, los péptidos fosforilados de referencia se sintetizan generalmente con la misma secuencia que algunos de los péptidos fosforilados que se desea cuantificar pero que están enriquecidos en isótopos pesados estables de carbono y nitrógeno. Los péptidos se sintetizan generalmente usando química en fase sólida en la que se añade un aminoácido a la vez para formar una cadena de aminoácidos o polipéptido. Generalmente, tales péptidos están enriquecidos en  $^{13}\text{C}$  y  $^{15}\text{N}$  que sustituyen a los  $^{12}\text{C}$  y  $^{14}\text{N}$  comunes. Este enriquecimiento da como resultado que los péptidos fosforilados de referencia sean aproximadamente de 6 a 10 daltons más pesados que los péptidos fosforilados endógenos con la misma secuencia para que puedan distinguirse usando un espectrómetro de masas.
- 10 Cuando la enzima modificadora de proteínas es una proteína acetiltransferasa y se están cuantificando péptidos acetilados, los péptidos modificados de referencia son péptidos acetilados de referencia. Dichos péptidos acetilados de referencia son generalmente péptidos sintéticos que contienen aminoácidos acetilados.
- 15 Los péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) se añaden generalmente en una cantidad conocida en cada una de las muestras a comparar. Las señales de los péptidos modificados endógenos (generalmente péptidos fosforilados) se normalizan a la señal de los péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) en el análisis posterior.
- 20 La etapa (b) de este método comprende además enriquecer péptidos modificados (generalmente péptidos fosforilados) a partir de la mezcla de péptidos y péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) obtenidos en la etapa (b) para producir una mezcla de péptidos modificados enriquecidos (generalmente péptidos fosforilados). Esta etapa adicional da como resultado una mezcla única de péptidos modificados enriquecidos (generalmente péptidos fosforilados) por muestra. En este contexto, la etapa (c) comprende, de este modo, llevar a cabo una espectrometría de masas (EM) en la mezcla de péptidos modificados enriquecidos (generalmente péptidos fosforilados) para obtener datos relacionados con los péptidos en la muestra. En este contexto, la etapa (b) generalmente da como resultado una mezcla de péptidos modificados enriquecidos (generalmente péptidos fosforilados).
- 25 La etapa de enriquecer péptidos modificados (generalmente péptidos fosforilados) se lleva a cabo generalmente usando cromatografía. La cromatografía puede ser cromatografía de afinidad de iones metálicos inmovilizados (IMAC), cromatografía de dióxido de titanio ( $\text{TiO}_2$ ) y/o cromatografía de dióxido de circonio ( $\text{ZrO}_2$ ). Generalmente, la cromatografía es cromatografía IMAC y  $\text{TiO}_2$ .
- 30 Alternativamente, la etapa de enriquecer péptidos modificados (generalmente péptidos fosforilados) se lleva a cabo utilizando métodos basados en anticuerpos.
- 35 Cuando la enzima modificadora de proteínas es una proteína quinasa y los péptidos que se van a cuantificar son péptidos fosforilados, los anticuerpos con afinidad a aminoácidos fosforilados tales como la tirosina, treonina, serina o histidina pueden unirse (inmovilizarse) a una matriz sólida. Los péptidos fosforilados se enriquecen por la capacidad de estos anticuerpos para unirse específicamente a los péptidos fosforilados. Los péptidos no fosforilados se lavan después mientras los péptidos fosforilados se retienen en las matrices recubiertas de anticuerpos. La elución de péptidos fosforilados a partir del anticuerpo inmovilizado se lleva a cabo generalmente usando disolventes de pH bajo o por cualquier otro método adecuado que desnaturalice la interacción entre el anticuerpo y los péptidos fosforilados.
- 40 Cuando la enzima modificadora de proteínas es una proteína acetiltransferasa, y los péptidos que se van a cuantificar son péptidos acetilados, los péptidos acetilados se enriquecen mediante el uso de anticuerpos específicos contra restos de aminoácidos acetilados. Dichos anticuerpos están unidos a una matriz sólida y después enriquecidos por la capacidad de los anticuerpos para unirse específicamente a restos de aminoácidos acetilados. Los péptidos no acetilados se lavan después mientras los péptidos acetilados se retienen en el anticuerpo inmovilizado.
- 45 En la etapa (c) de este método, la espectrometría de masas (EM) se lleva a cabo sobre la mezcla de péptidos y péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) obtenidos en la etapa (b) para obtener datos relacionados con los péptidos en la muestra. Generalmente, estos datos tienen la forma de un archivo de datos de EM para la muestra. Cuando la etapa (b) de este método comprende además enriquecer péptidos modificados (generalmente péptidos fosforilados) a partir de la mezcla de péptidos y péptidos modificados de referencia (generalmente péptidos fosforilados de referencia) obtenidos en la etapa (b) para producir una mezcla de péptidos modificados enriquecidos (generalmente péptidos fosforilados), la etapa (c) comprende llevar a cabo una espectrometría de masas (EM) en dicha mezcla de péptidos modificados enriquecidos (generalmente péptidos fosforilados) para obtener datos relacionados con los péptidos en la muestra, generalmente un archivo de datos de EM para la muestra. Generalmente, la espectrometría de masas es cromatografía líquida-espectrometría de masas (CL-EM). La etapa (c) por lo tanto generalmente da como resultado un archivo de datos de CL-EM (uno de cada muestra).
- 50 Los datos relacionados con los péptidos en la muestra generalmente comprenden la relación de masa a carga ( $m/z$ ), carga ( $z$ ) y/o tiempo de retención relativo de los péptidos.
- 55
- 60
- 65

En la etapa (d) de este método, los datos relacionados con los péptidos en la muestra (generalmente en forma de un archivo de datos de EM y más generalmente un archivo de datos CL-EM) se comparan con los datos en una base de datos de péptidos modificados (generalmente péptidos fosforilados) usando un programa informático. Por ejemplo, la relación masa a carga ( $m/z$ ), carga ( $z$ ) y tiempo de retención relativo de los péptidos en la muestra se comparan con la relación masa a carga ( $m/z$ ), carga ( $z$ ) y tiempo de retención relativo de los péptidos modificados (generalmente péptidos fosforilados) en la base de datos. Esto permite la identificación y cuantificación de cada péptido modificado (generalmente péptido fosforilado) en la muestra usando la base de datos de péptidos modificados (generalmente péptidos fosforilados).

Generalmente, el programa informático es el programa denominado PESCAL (Cutillas, P. R.; Vanhaesebroeck, B. Mol Cell Proteomics 6(9), 1560-73, 2007). PESCAL construye cromatogramas de iones extraídos (XIC, por sus siglas en inglés, es decir, un perfil de elución) para cada uno de los péptidos modificados (generalmente péptidos fosforilados) presentes en la base de datos en todas las muestras que se van a comparar. Esto se hace centrando el XIC en la  $m/z$  y el tiempo de retención del péptido previamente identificado a ser modificado (generalmente fosforilado) (es decir, presente en la base de datos construida en la primera etapa del procedimiento). PESCAL también considera la carga del péptido para ayudar en la asignación correcta de identidad. El programa también calcula la altura máxima y el área bajo la curva de cada XIC. Los datos se normalizan dividiendo la lectura de intensidad (áreas o alturas máximas) de cada péptido modificado (generalmente péptido fosforilado) que está siendo analizado entre las de los péptidos modificados de referencia (generalmente péptidos fosforilados de referencia).

En este contexto, la base de datos de péptidos modificados se recopila mediante un método que comprende las siguientes etapas:

- i obtener péptidos de una muestra;
- ii enriquecer péptidos modificados a partir de los péptidos obtenidos en la etapa i;
- iii realizar cromatografía líquida-espectrometría de masas en tándem (CL-EM/EM) en los péptidos modificados enriquecidos obtenidos en la etapa ii;
- iv comparar los péptidos modificados detectados en la etapa iii con una base de datos de referencia conocida para identificar los péptidos modificados; y
- v recopilar datos relacionados con los péptidos modificados identificados en la etapa iv en una base de datos.

La etapa i implica obtener péptidos de una muestra. Los péptidos se pueden obtener de la muestra usando cualquier método adecuado conocido en la técnica y como se describe en el presente documento. La muestra de la etapa i puede ser una tercera muestra.

La muestra generalmente es una muestra biológica y, por lo tanto, puede ser cualquier tipo de muestra obtenida de una fuente biológica, como se describe anteriormente. Generalmente, la muestra es una línea celular o un tejido.

Cuando la muestra utilizada en la etapa i es una línea celular, la muestra puede tratarse con un inhibidor antes de llevar a cabo la etapa i. El inhibidor puede ser cualquier tipo adecuado de inhibidor. Generalmente, cuando se están cuantificando péptidos fosforilados, el inhibidor es un inhibidor de fosfatasa. El tratamiento con inhibidores de fosfatasa aumenta la estequiometría de la fosforilación y da como resultado un mayor número de péptidos fosforilados que se pueden incluir en la base de datos. Además, los inhibidores de metil transferasa o acetil hidrolasa pueden usarse cuando el propósito es cuantificar péptidos metilados y acetilados, respectivamente.

La etapa i puede comprender:

- (1) lisar las células en una muestra;
- (2) extraer las proteínas de las células lisadas obtenidas en la etapa (1); y
- (3) escindir dichas proteínas en péptidos.

Estas etapas son las descritas anteriormente. Sin embargo, la etapa (3) se lleva a cabo generalmente usando el mismo método que en la etapa (a) descrita anteriormente.

En la etapa ii, los péptidos modificados (generalmente péptidos fosforilados) se enriquecen a partir de los péptidos obtenidos en la etapa i. La etapa ii da como resultado varias fracciones enriquecidas en péptidos modificados (generalmente péptidos fosforilados).

El enriquecimiento de péptidos modificados (generalmente péptidos fosforilados) en la etapa ii se lleva a cabo generalmente usando cromatografía multidimensional. En un caso, la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio catiónico fuerte (SCX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía de dióxido de titanio ( $TiO_2$ ). En otro caso, la cromatografía multidimensional se lleva a cabo usando cromatografía líquida de alta resolución de intercambio aniónico (SAX-HPLC), cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y cromatografía de dióxido de titanio ( $TiO_2$ ). En estos casos, las técnicas cromatográficas se llevan a cabo secuencialmente.

Alternativamente, el enriquecimiento de péptidos modificados (generalmente péptidos fosforilados) en la etapa ii se lleva a cabo utilizando métodos basados en anticuerpos, como se describe anteriormente.

5 En la etapa iii, La cromatografía líquida-espectrometría de masas en tándem (CL-EM/EM) se lleva a cabo sobre los péptidos modificados enriquecidos (generalmente péptidos fosforilados) obtenidos en la etapa ii.

10 En la etapa iv, los péptidos modificados (generalmente péptidos fosforilados) detectados en la etapa iii se comparan con una base de datos de referencia conocida para identificar los péptidos modificados (generalmente péptidos fosforilados). Esta etapa generalmente se lleva a cabo utilizando un motor de búsqueda disponible comercialmente, tales como, aunque sin restricción, los motores de búsqueda MASCOT, ProteinProspector o Sequest.

15 En la etapa v, los datos relacionados con los péptidos modificados (generalmente péptidos fosforilados) identificados en la etapa iv se recopilan en una base de datos. Esta base de datos enumera todos los parámetros necesarios para la cuantificación de péptidos fosforilados en experimentos biológicos posteriores. Generalmente, los datos relacionados con los péptidos modificados (generalmente péptidos fosforilados) incluyen la identidad de los péptidos modificados (generalmente péptidos fosforilados), la relación masa a carga ( $m/z$ ), carga y/o tiempo de retención relativo. Esto permite que los datos relacionados con los péptidos en la muestra, generalmente la relación de masa a carga ( $m/z$ ), carga ( $z$ ) y tiempo de retención relativo de los péptidos en la muestra, se comparen con los valores de los péptidos modificados (generalmente péptidos fosforilados) en la base de datos y, por lo tanto, permite la identificación y cuantificación de los péptidos modificados (generalmente péptidos fosforilados) en la muestra.

20 En este contexto, no es necesario que la recopilación de la base de datos se realice de forma simultánea con el método. La recopilación de la base de datos puede llevarse a cabo por separado, antes de que la técnica TIQUAS se use en el método para cuantificar los sitios de modificación en un péptido modificado en la muestra.

25 La base de la técnica TIQUAS es la construcción de una base de datos de péptidos modificados (generalmente péptidos fosforilados) que se pueden detectar y cuantificar por CL-EM. Esta base de datos enumera todos los parámetros necesarios para la cuantificación de péptidos modificados (generalmente péptidos fosforilados) en experimentos biológicos posteriores, incluida la identidad del péptido modificado (péptido generalmente fosforilado), relación masa/carga ( $m/z$ ), carga y tiempo de retención relativo. La base de datos puede construirse enriqueciendo péptidos modificados (generalmente péptidos fosforilados) usando cromatografía multidimensional (como intercambio catiónico fuerte, IMAC y  $TiO_2$ ). Las fracciones de péptidos modificados enriquecidos (generalmente péptidos fosforilados) pueden analizarse después por CL-EM/EM para la identificación de péptidos modificados (generalmente péptidos fosforilados).

30 El programa informático llamado PESCAL (Cutillas y Vanhaesebroeck, *Molecular & Cellular Proteomics* 6, 1560-1573 (2007)) automatiza la cuantificación de cada uno de los péptidos modificados (generalmente péptidos fosforilados) enumerados en la base de datos en ejecuciones de CL-EM de péptidos modificados (péptidos generalmente fosforilados) tomadas de experimentos biológicos. Para estos experimentos biológicos, las proteínas en los lisados celulares se digieren usando tripsina u otras proteasas adecuadas. Los patrones internos de péptidos (tales como fosfopéptidos), que son péptidos modificados de referencia (generalmente péptidos fosforilados de referencia), se añaden en cantidades conocidas en todas las muestras a comparar. Los péptidos modificados (generalmente péptidos fosforilados) en la mezcla de péptidos resultante se enriquecen usando una etapa de extracción IMAC o  $TiO_2$  sencilla de realizar. Los péptidos modificados enriquecidos (generalmente péptidos fosforilados) se analizan en una única ejecución de CL-EM de forma general pero sin restricción a aproximadamente 120 minutos (ciclo total). PESCAL construye después cromatogramas de iones extraídos (XIC, es decir, un perfil de elución) para cada uno de los péptidos modificados (generalmente péptidos fosforilados) presentes en la base de datos en todas las muestras que se van a comparar. El programa también calcula la altura máxima y el área bajo la curva de cada XIC. Los datos se normalizan dividiendo la lectura de intensidad (áreas o alturas máximas) de cada analito de péptido modificado (generalmente fosfopéptido) entre los del IS de péptido modificado (generalmente fosfopéptido).

35 Como alternativa al uso de la técnica TIQUAS, la cuantificación de modificaciones tales como la fosforilación también se puede llevar a cabo utilizando técnicas de EM que usan marcadores de isótopos para la cuantificación, tal como técnicas de marcado metabólico (por ejemplo, aminoácidos estables marcados con isótopos en cultivo, (SILAC); Olsen, J.V. *et al.* *Cell* 127, 635-648 (2006)) y de derivatización química (p.ej., iTRAQ (Ross, P. L.; *et al.* *Mol Cell Proteomics* 2004, 3, (12), 1154-69), ICAT (Gygi, S.P. *et al.* *Nat Biotechnol* 17, 994-999 (1999)), TMT (Dayon L *et al.*, *Anal Chem.* 2008 Apr 15;80(8):2921-31). En los métodos, las modificaciones de proteínas pueden cuantificarse con técnicas de CL-EM que miden las intensidades de los iones no fragmentados o con técnicas de CL-EM/EM que miden las intensidades de los iones fragmentados (como el Control de Reacción Seleccionada (SRM por sus siglas en inglés), también llamado control de reacciones múltiples (MRM por sus siglas en inglés)).

40 Los presentes inventores han ideado previamente una técnica para inferir sistemáticamente la activación de la ruta de proteínas quinasas a partir de datos de fosfoproteómica basados en EM. La técnica se denomina Análisis de Enriquecimiento de Sustrato de Quinasas (KSEA) y se describe en (Casado P, *et al.*, *supra*), así como en la solicitud de patente publicada como documento WO 2013/132075.

Como se describe en el documento WO 2013/132075, el método KSEA es un método para cuantificar la actividad de una enzima modificadora de proteínas en una muestra, que comprende:

5 (i) agrupar péptidos modificados de una primera muestra y péptidos modificados de una segunda muestra en un solo grupo de acuerdo con uno de los siguientes parámetros:

(a) péptidos modificados que tienen un sitio de modificación que está modificado por la misma enzima modificadora de proteínas; o

10 (b) péptidos modificados que tienen un sitio de modificación que es parte del mismo motivo de modificación;

(ii) calcular el enriquecimiento de los péptidos modificados de la primera muestra en comparación con los péptidos modificados de la segunda muestra del grupo; y

(iii) calcular la significación estadística de dicho enriquecimiento;

15 en donde un enriquecimiento estadísticamente significativo es indicativo de una enzima modificadora de proteínas que se activa en la primera muestra en comparación con la segunda muestra.

Para identificar péptidos modificados que tienen un sitio de modificación que está modificado por la misma enzima modificadora de proteínas, el método KSEA se basa en una recopilación de sitios de fosforilación conocidos de la literatura y de bases de datos tales como PhosphoSite (Hornbeck *et al.*, *Proteomics* 4, 1551 (Jun, 2004)) y PhosphoElm (Dinkel *et al.*, *Nucleic Acids Res* 39, D261 (enero, 2011)). Sin embargo, tales bases de datos no son completas. La presente invención permite la preparación de una base de datos de sitios de modificación de proteínas que puede usarse como entrada para el método KSEA sin depender de tales bases de datos. En particular, la base de datos que puede recopilarse usando el conjunto de datos de sitios de modificación producido por el método de la invención puede usarse como entrada para la etapa (i) (a) del método KSEA descrito anteriormente.

La presente divulgación se refiere a un método para crear un conjunto de datos de sitios de modificación, que comprende agrupar péptidos modificados de una primera muestra que se ha tratado con un primer modulador de una enzima modificadora de proteínas y péptidos modificados de una segunda muestra que se ha tratado con un segundo modulador de la misma enzima modificadora de proteínas en un solo grupo, de acuerdo con el efecto de dichos primer y segundo moduladores de dicha enzima modificadora de proteínas sobre dichos péptidos modificados, en donde dichos primer y segundo moduladores de dicha enzima modificadora de proteínas son diferentes.

El método puede comprender las etapas de

35 (a) tratar muestras (por ejemplo, células de una línea celular) con inhibidores contra un grupo de quinasas;

(b) medir los cambios resultantes en la fosforilación usando fosfoproteómica basada en EM;

(c) identificar sitios de fosforilación que se reducen significativamente en abundancia por al menos un inhibidor de quinasas;

40 (d) agrupar sitios de fosforilación basándose en el comportamiento en tratamiento con inhibidores contra la misma quinasa.

El resultado de este método son conjuntos de datos enriquecidos en sitios de fosforilación específicos de la quinasa prevista.

45 El conjunto de datos de los sitios de modificación producidos de acuerdo con el primer aspecto de la invención se puede usar después para preparar una base de datos.

Por consiguiente, en un segundo aspecto, la presente invención proporciona un método para preparar una base de datos, que comprende crear un conjunto de datos de sitios de modificación de acuerdo con un método del primer aspecto de la invención y recopilar dicho conjunto de datos en una base de datos.

La base de datos también puede incluir información que incluye, por ejemplo, las identidades de las proteínas que contienen los sitios de modificación, el tipo de modificación, el tipo de muestra en la que está presente el sitio de modificación y/o el modulador que aumenta o disminuye la modificación en el sitio de modificación. La base de datos se puede usar después como entrada para otros métodos para analizar datos tales como datos fosfoproteómicos, por ejemplo, el método KSEA ideado previamente por los presentes inventores, y se puede usar para identificar marcadores o biomarcadores de quinasas que son inhibidos por compuestos particulares o están asociados con fenotipos particulares de células en las que se encuentran los sitios de modificación.

60 De acuerdo con un tercer aspecto, la invención proporciona un método *in vitro* para diagnosticar una enfermedad que comprende:

(a) Crear un conjunto de datos de prueba de sitios de modificación:

65 i. Tratando una o más muestras de prueba de un sujeto con uno o más moduladores de una enzima

modificadora de proteínas,

ii. Identificando y/o cuantificando sitios de modificación en péptidos modificados en una o más muestras de prueba,

iii. Agrupando sitios de modificación en péptidos modificados de una o más muestras de prueba en un solo grupo, creando así el conjunto de datos de prueba de sitios de modificación, y

iv. opcionalmente recopilando el conjunto de datos de prueba en una base de datos de prueba,

(b) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el primer aspecto de la invención o comparar la base de datos de prueba con una base de datos preparada de acuerdo con el segundo aspecto de la invención,

(c) Encontrar una diferencia significativa entre el conjunto de datos de prueba o la base de datos de prueba y el conjunto de datos de control o la base de datos de control,

(d) Diagnosticar que el sujeto tiene la enfermedad si se encuentra una diferencia significativa entre el conjunto de datos o la base de datos en la etapa (c).

Las etapas (b) y (c) pueden realizarse por cualquier medio adecuado. La etapa (b) puede comprender el cálculo de una reducción o aumento en la abundancia de sitios de modificación en péptidos modificados de la una o más muestras de prueba en relación con el conjunto de datos de los sitios de modificación creado de acuerdo con el primer aspecto de la invención o comparar la base de datos de prueba con una base de datos preparada de acuerdo con el segundo aspecto de la invención. La reducción o aumento de la abundancia es generalmente una reducción o aumento estadísticamente significativo y puede calcularse usando cualquier método estadístico adecuado, por ejemplo, como se describe en el Ejemplo en el presente documento.

En algunas realizaciones, la enfermedad es cáncer.

De acuerdo con un cuarto aspecto, la invención proporciona un método para identificar uno o más biomarcadores para determinar la actividad de una enzima modificadora de proteínas que comprende crear un conjunto de datos de sitios de modificación de acuerdo con el método del primer aspecto de la invención o preparar una base de datos de acuerdo con el segundo aspecto de la invención, en donde los sitios de modificación en el único grupo se identifican como el uno o más biomarcadores para la actividad de una enzima modificadora de proteínas.

De acuerdo con un quinto aspecto, la invención proporciona un método *in vitro* para determinar que una sustancia de prueba es un modulador de una enzima modificadora de proteínas que comprende tratar una muestra de prueba con la sustancia de prueba, identificar y/o cuantificar sitios de modificación en péptidos modificados en la muestra de prueba, crear un conjunto de datos de prueba de sitios de modificación a partir de los sitios de modificación identificados y/o cuantificados en la muestra de prueba, que comprende además:

(a) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el método del primer aspecto de la invención y/o

(b) Recopilar dicho conjunto de datos de prueba en una base de datos de prueba y comparar la base de datos de prueba con una base de datos preparada de acuerdo con el segundo aspecto de la invención.

Características preferidas para el segundo aspecto y los aspectos posteriores de la invención son como para el primer aspecto *mutatis mutandis*.

La presente invención se describirá a continuación adicionalmente a modo de referencia a los siguientes Ejemplos que están presentes solo con fines ilustrativos. En los Ejemplos, Se hace referencia a una serie de Figuras en las que:

Fig. 1. Los datos de fosfoproteómica aclaran las relaciones entre los inhibidores de quinasas. (A) Diseño experimental. Inh, inhibidor. (B) Inhibidores de quinasas utilizados en el estudio. (C) Análisis del componente principal (PCA) de los 4.651 sitios de fosforilación cuya abundancia se redujo significativamente ( $P$  ajustado  $<0,05$ ) mediante al menos un tratamiento con inhibidor. Los números en superíndice representan los tratamientos con inhibidores como en (B). PC, componente principal. (D) Triángulo inferior: Coeficientes de correlación de Pearson entre cada uno de los tratamientos con inhibidores. Las relaciones conocidas de quinasa-quinasa se resaltan con recuadros blancos discontinuos. Triángulo superior:  $\log_2$  de los factores de proporción de alineaciones por pares de los 4.651 sitios de fosforilación para cada combinación de inhibidores. La línea en cada recuadro indica el modelo lineal formado entre las dos variables; la elipse en cada recuadro representa una desviación estándar de la media en ambas dimensiones. (E) Agrupación jerárquica no supervisada (métrica de distancia de correlación de Pearson) del  $\log_2$  de la media de las proporciones para péptidos que contienen motivos de fosforilación comunes representados en los 4.651 sitios de fosforilación filtrados.

Fig. 2. Inferencia de una topología de red de señalización de quinasas a partir de datos de fosfoproteómica. (A)  $\log_2$  de los factores de proporción de sitios de fosforilación (frente a DMSO control) para los dos inhibidores de Akt; MK-2206 e inhibidor de Akt VIII. Las líneas punteadas representan los umbrales para la identificación de CTAM. Los sitios de fosforilación indicados usando una impresión más oscura en el cuadrante negativo del gráfico (abajo a la izquierda) representan los marcadores de actividad compuesto-diana (CTAM) identificados para Akt.

Puntos de datos sombreados oscuros debajo de la línea punteada horizontal inferior y a la izquierda de la línea punteada más a la izquierda,  $FR \leq -1,0$ ,  $P_{ajust.} \leq 0,1$  para ambos inhibidores; puntos de datos sombreados oscuros debajo de la línea punteada horizontal superior y a la izquierda de la línea punteada más a la derecha,  $FR \leq -0,75$ ,  $P_{ajust.} \leq 0,05$  para ambos inhibidores; también se muestran datos de umbrales mezclados entre inhibidores. (B) 610 sitios de fosforilación identificados como que son marcadores de actividad compuesto-diana para al menos una quinasa. (C) Ejemplos ilustrativos de sitios de fosforilación dispuestos en grupos de CTAM. \*  $FR \leq -0,75$ ,  $P_{ajust.} \leq 0,05$ ; \*\*  $FR \leq -1,0$ ,  $P_{ajust.} \leq 0,1$ ; \*\*\*  $FR \leq -1,0$ ,  $P_{ajust.} \leq 0,01$ . Sombreado como en (B). (D) Un grafo de red bipartita no dirigida que muestra todos los marcadores de actividad identificados. El diseño del grafo de red se basa en un algoritmo de dibujo dirigido por fuerza. Los nodos grandes representan las quinastas dirigidas en el experimento (asociadas a MAPK (EGFR, MEK, ERK); asociadas a PI3K/mTOR/p70S6K (PI3K, mTOR, p70S6K); asociación mixta (Akt, CAMK2, ROCK, PKC)). Los nodos más pequeños representan sitios de fosforilación individuales. Los bordes grises indican si el sitio de fosforilación es un CTAM de la quinasa a la que está conectada.

Fig. 3. Cinética del comportamiento del grupo de CTAM tras la estimulación con factor de crecimiento. (A) Perfiles temporales de cada uno de los grupos de CTAM seleccionados en células tratadas con factores de crecimiento (EGF: paneles de la izquierda; IGF-1: paneles de la derecha;  $m$  = número de sitios de fosforilación cuantificados en el grupo de CTAM nombrado). Los puntos de datos representan la media  $\pm$  desviación estándar (DE). \*\*\*  $P < 0,001$ ; \*\*  $P < 0,01$ ; \*  $P < 0,05$ . (B) Perfiles temporales para cada uno de los grupos de CTAM representados en el conjunto de datos (con  $m \geq 2$ ). Los tamaños de los puntos de datos son proporcionales al  $\log_2$  de la media del factor de proporción (frente a  $t = 0$  min) y están sombreados de acuerdo con la significación estadística del enriquecimiento. La agrupación jerárquica no supervisada se basó en la métrica de distancia euclidiana. (C) Distribución del coeficiente de variaciones para cada uno de los grupos de CTAM cuantificados en cada uno de los puntos de tiempo. Las líneas verticales indican la media.

Fig. 4. Evolución de la actividad de la red de señalización en células resistentes a mTORC1/2 e inhibidores de PI3K. Crecimiento de cultivos de células MCF7 precursoras y resistentes a fármacos en presencia de GDC-0941(A) o KU-0063794(B). Los puntos de datos representan la media  $\pm$  DE. (C) Perfiles del grupo de CTAM (con  $m \geq 2$ ) para cada una de las líneas celulares resistentes (res.) en comparación con la línea celular precursora (par. por sus siglas en inglés). Los tamaños de los puntos representan  $\log_2$  de la media del factor de proporción de cada grupo de CTAM con respecto a la línea celular precursora, normalizado a la abundancia de proteína no modificada. El sombreado representa la significación del enriquecimiento. La agrupación jerárquica de los grupos de CTAM se basó en la métrica de distancia euclidiana. (D) PCA de los datos fosfoproteómicos que se muestran en C. (E) PCA de los datos de viabilidad celular (medidos por MTS) en función del tratamiento con un grupo de inhibidores de quinastas (que se muestra en el Apéndice detallado, Fig. 14A). (F) Como en (E), sin embargo, la viabilidad celular se midió por tinción con cristal violeta (se muestra en el Apéndice detallado, Fig. 14B).

Fig. 5. Resumen de datos para los fosfoproteomas tratados con inhibidores de quinasa que definen la red. Este es un análisis de control de calidad de los datos mostrados en las Fig. 1 y 2. (A) Histograma que muestra la distribución de los valores  $q$  dentro de la base de datos de iones de fosfopéptidos (13.405 entradas únicas). (B) Diagramas de cajas que demuestran el  $\log_2$  de las distribuciones de intensidad de iones de fosfopéptidos para cada normalización pre- (panel superior) y post-cuantil de la muestra (panel inferior). (C) Ejemplo ilustrativo de una alineación del tiempo de retención patrón del péptido enolasa ( $t_R$ ) entre dos muestras para predecir y contar los cambios en el tiempo de retención. (D) Histograma que muestra la distribución de los coeficientes de correlación de Pearson ( $r$ ) para modelos lineales formados entre cada alineación patrón de péptidos por pares. (E) Diagrama lineal que demuestra la estabilidad de los tiempos de retención patrón del péptido enolasa a través de la matriz de muestras. El sombreado de cada línea es específico de un solo ion de péptido de enolasa (como se indica en la leyenda).

Fig. 6. Significación estadística de las correlaciones entre los tratamientos con inhibidores. Mapa de calor que demuestra la significación de la correlación entre cada tratamiento con inhibidor (basado en los 4.651 iones de fosfopéptidos que se encuentran significativamente regulados negativamente por al menos un par de inhibidores, como en la Fig. 1C). La significación se expresa como un valor de  $P$  (calculado a partir de  $r$  de Pearson y ajustado para múltiples pruebas usando el procedimiento Holm) y se codifica sombreado en consecuencia. Las relaciones conocidas de quinasa-quinasa se resaltan con recuadros blancos discontinuos como en la Fig. 1C.

Fig. 7. Número de CTAM identificados en función de umbrales estadísticos seleccionados. Las barras representan el número de CTAM identificadas en las condiciones de umbral indicadas.  $FR = \log_2$  del factor de proporción;  $P_{ajust.}$  = Valor de  $P$  ajustado por Benjamini-Hochberg;  $\delta$  = puntuación delta de Mascot ( $\delta$ ).

Fig. 8. Ejemplo de datos sin procesar de EM. Los datos sin procesar de EM soportan la identificación de AKT1S como miembro del grupo de CTAM Akt-mTOR-PI3K. Los datos sin procesar de EM extrajeron cromatogramas iónicos (XIC) para el fosfopéptido que representa AKT1S (S183) en  $m/z$  698,3588 ( $\pm 7$  ppm). Las líneas representan los perfiles de elución temporales del primer, segundo y tercer isótopos como se describe en la leyenda (en cada caso, el pico más alto representa el primer isótopo, el segundo pico más alto representa el segundo isótopo y el pico más bajo representa el tercer isótopo). Las líneas punteadas representan el tiempo de retención previsto para el fosfopéptido en cada muestra. Akt1, inhibidor de Akt VIII; Akt2, MK-2206; mTOR1, KU-0063794; mTOR2, Torin-1; PI3K1, GDC-0941; PI3K2, PI-103.

Fig. 9. Aleatorización de la topología de la red definida. (A) Red de señalización de quinasas definida empíricamente como se muestra en la Fig. 2D. (B) Optimización del número de etapas de cambio necesarias para producir redes menos similares a la original. Línea vertical, el número de etapas de cambio utilizadas en posteriores intentos de aleatorización de la red. (C) Un ejemplo ilustrativo de una de las 10.000 redes aleatorizadas (producidas usando 50.000 etapas de cambio). (D) Distribución de solapamientos de CTAM entre las quinasas indicadas en las redes aleatorizadas. Línea curva, función de densidad de probabilidad de los solapamientos observados en las 10.000 redes aleatorizadas (como en C); línea vertical, solapamiento observado en la red real (como en A).

Fig. 10. Resumen de datos para la estimulación de células MCF7 con factores de crecimiento. Este es un análisis de control de calidad de los datos mostrados en la Fig. 3. (A) Histograma que muestra la distribución de los valores  $q$  dentro de la base de datos de iones de fosfopéptidos (14.559 entradas únicas). (B) Diagramas de cajas que demuestran el  $\log_2$  de las distribuciones de intensidad de iones de fosfopéptidos para cada normalización pre- (panel superior) y post-cuantil de la muestra (panel inferior). (C) Ejemplo ilustrativo de una alineación del tiempo de retención patrón ( $t_R$ ) del péptido enolasa entre dos muestras, una del primer experimento que define la red (Fig. 5 - expt 1) y otra del experimento de estimulación del factor de crecimiento (expt 2), para predecir y contar los cambios en el tiempo de retención. (D) Histograma que muestra la distribución de los coeficientes de correlación de Pearson ( $r$ ) para modelos lineales formados entre cada alineación patrón de péptidos por pares, tanto en los experimentos de definición de redes como en los de factores de crecimiento. (E) Estabilidad de los tiempos de retención patrón del péptido enolasa en toda la matriz de muestras. El sombreado de cada línea es específico de un solo ion de péptido de enolasa (como se indica en la leyenda).

Fig. 11. Ejemplos ilustrativos de las dinámicas de los sitios de fosforilación cadena abajo de EGFR e IGF-1R. Mapa de calor que muestra el  $\log_2$  de los factores de proporción (frente a  $t = 0$ ) para varios sitios de fosforilación que se sabe que están cadena abajo de EGFR e IGF-1R después de la estimulación con cualquiera de los factores de crecimiento. Los sitios de fosforilación que poseen una puntuación delta de Mascot  $\geq 10$  se presentan como el nombre de la proteína, carga peptídica ( $z$ )  $\pm$  cualquier otra modificación peptídica. Aquellos que poseen una puntuación delta de Mascot  $\leq 10$  se representan como el nombre de la proteína, los restos entre los que reside el sitio de fosforilación, la carga peptídica y la modificación prevista. Los asteriscos indican significación estadística (valor de  $P$  ajustado) como se indica en la leyenda. (B) transferencias Western para sitios de fosforilación sensibles a EGF/IGF conocidos que activan las rutas Akt-mTOR-PI3K y MAPK.

Fig. 12. Resumen de datos para la comparación cuantitativa de fosfoproteómica y proteómica de líneas celulares resistentes y células precursoras. Este es un análisis de control de calidad de los datos mostrados en la Fig. 4. (A) Histograma que muestra la distribución de los valores  $q$  dentro de la base de datos de iones de fosfopéptidos (15.325 entradas únicas). (B) Diagramas de cajas que demuestran el  $\log_2$  de las distribuciones de intensidad de iones de fosfopéptidos para cada normalización pre- (panel superior) y post-cuantil de la muestra (panel inferior). (C) Ejemplo ilustrativo de una alineación del tiempo de retención patrón ( $t_R$ ) del péptido enolasa entre dos muestras, una del primer experimento que define la red (Fig. 5 - expt 1) y otra del experimento de comparación de la línea celular resistente (expt 3), para predecir y contar los cambios en el tiempo de retención. (D) Histograma que muestra la distribución de los coeficientes de correlación de Pearson ( $r$ ) para modelos lineales formados entre cada alineación patrón de péptidos por pares, tanto en los experimentos de definición de redes como en los de líneas celulares resistentes. (E) Estabilidad de los tiempos de retención patrón del péptido enolasa en toda la matriz de muestras. El sombreado de cada línea es específico de un solo ion de péptido de enolasa (como se indica en la leyenda).

Fig. 13. Ejemplos ilustrativos de sitios de fosforilación indicados cadena abajo de PI3K y mTORC1/2 en líneas celulares resistentes tratadas con inhibidores. Mapa de calor que demuestra el  $\log_2$  de las proporciones estequiométricas ( $\log_2$  del factor de proporción del sitio de fosforilación/ $\log_2$  del factor de proporción de la proteína) frente a la línea celular precursora de varios sitios de fosforilación conocidos como cadena abajo de PI3K y mTORC1/2. Los sitios de fosforilación que poseen una puntuación delta de Mascot  $\geq 10$  se presentan como el nombre de la proteína, carga peptídica ( $z$ )  $\pm$  cualquier otra modificación peptídica. Aquellos que poseen una puntuación delta de Mascot  $\leq 10$  se representan como el nombre de la proteína, los restos entre los que reside el sitio de fosforilación, la carga peptídica y la modificación prevista. Los asteriscos indican significación estadística (valor de  $P$  ajustado) como se indica en la leyenda. (B) Transferencias Western para marcadores conocidos de la actividad de la ruta Akt-mTOR-PI3K.

Fig. 14. Datos de viabilidad celular para las líneas celulares resistentes y precursoras tratadas con un grupo de inhibidores de quinasas. Las barras representan el  $\log_2$  de la media del factor de proporción de la viabilidad celular (frente a DMSO) para cada inhibidor de quinasa solo o en combinación en cada una de las líneas celulares resistentes (MCF7-G, barras sólidas, sombreado oscuro, paneles superiores (G1, G2, G3); MCF7-K, barras sólidas, sombreado claro, paneles inferiores (K1, K2, K3); MCF7-P, barras blancas punteadas (G1, G2, G3, K1, K2, K3)) como se determina por el ensayo MTS (A) y los ensayos con cristal violeta (B). Las barras de error representan la desviación estándar de la media en cada caso ( $n = 3$ ). Estos datos fueron la entrada para el PCA presentada en la Fig. 4E. PI3K-i = GDC-0941 1  $\mu$ M; mTOR-i = KU-0063794 1  $\mu$ M; Akt-i = MK-2206 1  $\mu$ M; CAMK2-i = KN-93 5  $\mu$ M; EGFR-i = PD-153035 1  $\mu$ M; MEK-i = GSK-1120212 0,5  $\mu$ M.

## Ejemplos

### Métodos

5 Líneas celulares. La línea celular precursora MCF7 y las células resistentes MCF7-G1, MCF7-G2, MCF7-G3, MCF7-K1, MCF7-K2 y MCF7-K3, se cultivaron en DMEM (suplementado con suero fetal bovino al 10 % y 100 U.ml<sup>-1</sup> de penicilina/estreptomicina) a 37 °C en una atmósfera humidificada de CO<sub>2</sub> al 5 %. Después del tratamiento, como se indica en el texto, las células se lisaron en tampón de lisis de urea y las proteínas se digirieron con tripsina.

10 Fosfoproteómica basada en espectrometría de masas. Los péptidos fosforilados se enriquecieron usando perlas de TiO<sub>2</sub> (GL Sciences) de manera similar a la descrita anteriormente (Casado *et al.* (2013) *Sci. Signal.* 6(268):rs6., Montoya *et al.* (2011) *Methods* 54 (4): 370-378) con algunas modificaciones, y analizados por CL-EM/EM usando un espectrómetro de masas LTQ-Orbitrap. Los péptidos se identificaron por medios de búsquedas de Mascot contra la base de datos de proteínas humanas SwissProt. La cuantificación de péptidos se logró usando Piscal como se describe previamente (Casado *et al.* (2013) *Sci. Signal.* 6(268):rs6.).

15 Análisis estadístico. Después de la normalización cuantil de los datos (Bolstad *et al.* (2003) *Bioinformatics* 19 (2): 185-193), la magnitud y la significación estadística de las diferencias entre las condiciones se calcularon utilizando una reducción por Bayes empírico de las desviaciones estándar (Smyth (2004) *Stat. Appl. Genet. Molec. Biol.* 3(1):Article3) utilizando el paquete *limma* dentro del entorno informático R (Smyth (2005) *Limma: linear models for microarray data* (Springer New York); R Core Team (2013) *R: A language and environment for statistical computing* (R Foundation for Statistical Computing)). La abundancia de CTAM se monitorizó sistemáticamente utilizando KSEA (análisis de enriquecimiento de sustrato de quinasas), como se describe anteriormente (Casado *et al.* (2013) *Sci. Signal.* 6(268):rs6; Casado *et al.* (2013) *Genome Biol.* 14(4):R37; Montoya *et al.* (2011) *Methods* 54(4):370-378; Alcolea *et al.* (2012) *Mol. Cell. Proteomics* 11(8):453-466).

Se proporciona una descripción más detallada de estos métodos en Materiales y Métodos Detallados, más adelante.

### Materiales y Métodos Detallados

30 Reactivos. El inhibidor de Akt VIII, KU-0063794, GDC-0941, PI-103 e Y-27632 se obtuvieron de Chemdea. KN-93, KN-62, PD-168393, PD-153035, inhibidor de ERK, inhibidor de ERK II, U0126, PF-4708671, DG2, G6-6976, bisindolil-maleimida-1, ácido ocaidaico, ácido cantárico y H-1152 se obtuvieron de Calbiochem (Merck). GSK-1120212 y MK-2206 se obtuvieron de Selleckchem. Torin-1 se obtuvo de Axon Medchem. El EGF humano recombinante (AF-100-15) y el IGF-1 (100-11) se adquirieron de Peprotech.

35 Líneas celulares. Las células MCF7 (identidad confirmada por análisis de genotipo) se mantuvieron de forma rutinaria en DMEM (suplementado con suero fetal bovino al 10 % y 100 U.ml<sup>-1</sup> de penicilina/estreptomicina) a 37 °C en una atmósfera humidificada con CO<sub>2</sub> al 5 %. Líneas de células MCF7 resistentes, derivadas de las células precursoras originales, se obtuvieron aumentando gradualmente las concentraciones de GDC-0941 o KU-0063794 (comenzando a 100 nM hasta un máximo de 1 µM) durante un período de aproximadamente seis meses. Los clones resultantes (MCF7-G1:G3 y MCF7-K1:K3) se mantuvieron después de forma rutinaria como anteriormente en presencia de GDC-0941 o KU-0063794 (en DMSO) 1 para células MCF7-G y MCF7-K, respectivamente.

45 Lisis celular y digestión de proteínas. Las células se fraccionaron y sembraron en 4 x 10<sup>5</sup> células.placa<sup>-1</sup>, 72 horas antes del experimento. El medio en cada placa se reemplazó 24 horas antes del experimento. Cada experimento se realizó por duplicado biológico. Para los experimentos de definición de redes, las células se trataron con cada uno de los inhibidores o vehículos (DMSO) durante 1 hora (Tabla 1). Para los experimentos de estimulación del factor de crecimiento, las células se privaron de alimento durante 24 horas y posteriormente se trataron con 50 ng.ml<sup>-1</sup> de EGF o IGF durante 0, 5, 10, 30 o 60 minutos. Para los experimentos de las líneas celulares resistentes, no se realizaron tratamientos adicionales. Para los experimentos de reconexión en las líneas celulares MCF7-P y -G1:3, las células se privaron de alimento durante 24 h y posteriormente se trataron con 50 ng.ml<sup>-1</sup> de EGF durante 0, 30, 60 o 120 minutos. Después de los tratamientos, las células se lavaron tres veces con solución salina tamponada con fosfato enfriada con hielo suplementada con Na<sub>3</sub>VO<sub>4</sub> 1 mM y NaF 1 mM, y se lisaron con tampón de lisis de urea [urea 8 M en HEPES 20 mM (pH 8,0), suplementado con Na<sub>3</sub>VO<sub>4</sub> 1 mM, NaF 1 mM, Na<sub>2</sub>P<sub>2</sub>H<sub>2</sub>O<sub>7</sub> 1 mM y β-glicerol fosfato 1 mM]. Los lisados se homogeneizaron adicionalmente mediante sonicación con sonda (tres pulsos de 10 s) y el material insoluble se eliminó mediante centrifugación. La concentración de proteínas se estimó usando el ensayo de Bradford o Smith. Después de normalizar cada condición a una concentración de proteínas común (250 o 500 µg), cada muestra se redujo y alquiló por incubación secuencial con ditiotretol 10 mM y yodoacetamida 16,6 mM durante 30 minutos a temperatura ambiente, en la oscuridad. Para la digestión de proteínas, se redujo la concentración de urea a 2 M mediante la adición de HEPES 20 M (pH 8,0). Después se añadió tosil-lisina clorometilcetona (TLCK)-tripsina inmovilizada [20 p-toluenosulfonil-L-arginina metil éster (TAME) unidades.mg<sup>-1</sup>], y las muestras se incubaron durante la noche a 37 °C. Las perlas de tripsina se eliminaron por centrifugación y las soluciones de péptidos resultantes se desalaron con cartuchos Oasis-HLB de 30 mg (Waters, Manchester, GB), utilizando un colector de vacío. En resumen, los cartuchos Oasis se acondicionaron con acetonitrilo (ACN) y se equilibraron con solución de lavado (TFA al 0,1 %, ACN al 1 %). Los péptidos se cargaron en los cartuchos y se lavaron con 1 ml de solución de lavado. Por último, los

péptidos se eluyeron con tampón de ácido glicólico 1 (ácido glicólico 1 M, TFA al 5 %, ACN al 50 %). Al analizar la proteína total, se tomaron y procesaron alícuotas de los lisados de las líneas celulares resistentes originales y precursoras como se indica anteriormente, con la excepción de que las muestras se eluyeron de los cartuchos OASIS con ACN al 60 % (H<sub>2</sub>O al 40 %), se secaron usando un SpeedVac y se almacenaron a -80 °C hasta su posterior análisis.

Enriquecimiento de fosfopéptidos. Los péptidos fosforilados se enriquecieron usando TiO<sub>2</sub> (GL Sciences) de manera similar a la descrita anteriormente (1, 2) con algunas modificaciones descritas en los Métodos Detallados. En resumen, los eluyentes peptídicos se normalizaron con tampón de ácido glicólico 2 (ácido glicólico 1 M, TFA al 5 %, ACN al 80 %) e incubaron con 50 µl de TiO<sub>2</sub> (una suspensión al 50 % en TFA al 1 %), durante 5 minutos a temperatura ambiente.

Las perlas se sedimentaron por centrifugación, se retiró el 80% del sobrenadante y se almacenó en hielo. La solución restante se usó para resuspender las perlas, y las perlas se empaquetaron en cabezales giratorios prelavados, vacíos, filtrados con PE (Glygen, MD, EE. UU.) mediante centrifugación. Las perlas residuales se resuspendieron con otro tampón de ácido glicólico 2 y se empaquetaron en los cabezales. La solución de péptidos restante se retiró del hielo y se lavó sobre las perlas de TiO<sub>2</sub> empaquetadas por centrifugación. Los cabezales empaquetados se lavaron secuencialmente con tampón de ácido glicólico 2, tampón de acetato de amonio (acetato de amonio 100 mM; ACN al 25 %) y ACN al 10 %, repitiéndose esto último por triplicado. Los fosfopéptidos se eluyeron por centrifugación con cuatro lavados secuenciales de NH<sub>4</sub>OH al 5 % en ACN al 10 %. Las soluciones de fosfopéptidos resultantes se congelaron rápidamente, se secaron en un SpeedVac y se almacenaron a -80 °C.

Análisis de fosfoproteómica por CL-EM/EM. Los extractos de fosfopéptidos secos se resuspendieron en 20 µl de TFA al 0,1 % (ACN al 5 %) que contenía 20 fmol.µl<sup>-1</sup> de digestión de enolasa (Waters, Manchester, GB). Para cada repetición técnica, se cargaron 3,0 µl de cada muestra en un sistema LC Dionex Ultimate nRSLC 3000 (Thermo Fisher Scientific) acoplado en línea a un espectrómetro de masas LTQ-Orbitrap-Velos (Thermo Fisher Scientific). Las muestras se separaron sobre un gradiente lineal de 85 minutos entre ACN al 5 y 35 % en una columna Acclaim PepMap RSLC (25 cm x 75 µm, 2 µm, 100 Å) y se seleccionaron los siete iones principales con carga múltiple más intensos en cada exploración de EM<sup>1</sup> para la fragmentación por disociación inducida por colisión (con activación de etapas múltiples habilitada). La resolución de la EM<sup>1</sup> se estableció en 30.000 FWHM. Cada muestra se ejecutó por triplicado.

Análisis de proteómica por CL-EM/EM. Los extractos de péptidos secos se resuspendieron a una concentración final de 0,5 µg.µl<sup>-1</sup> en TFA al 0,1 %, ACN al 3 %. Para cada repetición técnica, se cargaron 4,0 µl de cada muestra en el sistema CL-EM/EM descrito anteriormente. Las muestras se separaron sobre un gradiente lineal de 120 minutos de ACN entre el 3 y el 32 % y se seleccionaron los diez iones principales con carga múltiple más intensos en cada exploración EM<sup>1</sup> para la fragmentación por disociación inducida por colisión. La resolución de la EM<sup>1</sup> se estableció en 30.000 FWHM. Cada muestra se ejecutó por duplicado.

Identificación y cuantificación de fosfopéptidos. La identificación de los péptidos se realizó haciendo coincidir los datos de EM/EM desisotopados con las bases de datos de proteínas humanas Uniprot-Swissprot (versión de octubre de 2012, que contiene 20.233 entradas), utilizando el servidor Mascot versión 2.3. La versión 2 del Destilador Mascot se utilizó para generar listas de máximos en el formato genérico mascot. Las muestras también se buscaron en una versión codificada de la base de datos anterior. Las tolerancias de masa se establecieron en 10 ppm y 600 mmu para los iones precursores y fragmentos respectivamente. Para los experimentos de fosfoproteómica, las modificaciones variables permitidas fueron fosfo-Ser, fosfo-Thr, fosfo-Tyr, piro-Glu (N-terminal) y oxidación-Met. Los fosfopéptidos identificados de cada una de las muestras se cotejaron, se curaron y se calculó una tasa de descubrimiento falso (en comparación con la base de datos codificada) utilizando scripts de desarrollo interno. Los iones de fosfopéptidos únicos con FDR <5 % se incluyeron después en los análisis posteriores. Usando estos ajustes > 95 % de los péptidos tenía una probabilidad de FDR <1 % (Fig. 5). Las bases de datos de los experimentos de factores de crecimiento y líneas celulares resistentes se unieron con la base de datos de descubrimiento de red original para garantizar la adición de los marcadores de actividad identificados. La cuantificación de péptidos se realizó según lo descrito anteriormente por el grupo de la presente invención (1-4) y otros (5, 6). En resumen, se utilizó el software Pescal (escrito en Python v2.7) para obtener las alturas máximas de los cromatogramas de iones extraídos de cada uno de los iones de fosfopéptidos en la base de datos, en todas las muestras a comparar. Los tiempos de retención de cada ion de fosfopéptido en cada muestra se predijeron alineando los péptidos patrón de enolasa añadidos a cada muestra, utilizando un algoritmo de modelado lineal interno y la posterior integración de los máximos cromatográficos obtenidos de los cromatogramas de iones extraídos para cada fosfopéptido en cada muestra. Este enfoque es similar al informado para la alineación de los tiempos de retención en experimentos de monitorización de reacción seleccionados multiplexados (7). Las tolerancias de masa a carga (m/z) y tiempo de retención (t<sub>R</sub>) se establecieron en 7 ppm y 1,5 min, respectivamente. Los datos de proteómica se procesaron como anteriormente; sin embargo, la identidad de las proteínas se infirió solo para aquellos que poseían ≥ 2 péptidos con puntuación de péptido Mascot ≥ 20 y una puntuación de proteína Mascot ≥ 60.

Análisis estadístico y visualización de datos. Después de la normalización cuantil de los datos (8), se realizó un análisis estadístico consistente para evaluar la magnitud y significación de los cambios de fosforilación utilizando una estrategia de modelado lineal y la reducción por Bayes empírico de las desviaciones estándar (9). Los valores de P resultantes

se corrigieron para múltiples pruebas usando el procedimiento Benjamini-Hochberg. Todos los análisis descritos se realizaron utilizando el paquete *limma* (v3.16.2) dentro del entorno informático estadístico R (v3.0.0) (10, 11). Los conjuntos de datos se visualizaron usando una combinación de paquetes R individuales, concretamente: *ggplot2*, *gplots*, *reshape2*, *igraph*, y *RCytoscape*, (12-16). Los gráficos de red se construyeron dentro del paquete de software Cytoscape (v2.8.3) (17) y se aleatorizaron usando el paquete *BiRewire* R (18). El  $\log_2$  de las proporciones estequiométricas se calculó restando el  $\log_2$  de la proporción calculada (frente al control) de la proteína total del  $\log_2$  del factor de proporción (frente al control) del sitio de fosforilación individual.

Análisis de enriquecimiento de sustrato de quinasas y análisis de motivos. La abundancia de cada uno de los marcadores identificados se monitorizó sistemáticamente utilizando KSEA (análisis de enriquecimiento de sustrato de quinasas), como se describe anteriormente (1). En resumen, los iones de fosfopéptidos identificados dentro de los experimentos de factor de crecimiento y de líneas celulares resistentes que previamente se identificaron como CTAM se extrajeron sistemáticamente del conjunto de datos. La media del  $\log_2$  de los factores de proporción (frente a control) de los sitios de fosforilación que representan cada grupo de CTAM se calcularon junto con la desviación estándar ( $\delta$ ) para cada media. La significación del enriquecimiento se determinó después utilizando un enfoque basado en la puntuación Z y se expresó como valores de P, usando la siguiente fórmula:  $Z = (S_m - \mu_t) \cdot m^{1/2} / \delta_t$ ; donde  $S_m$  = media del  $\log_2$  de los factores de proporción del grupo de CTAM;  $\mu_t$  = media del  $\log_2$  de las proporciones del conjunto de datos total;  $m$  = número de fosfopéptidos individuales dentro del grupo de CTAM; y  $\delta_t$  = desviación estándar del  $\log_2$  de las proporciones del conjunto de datos total (frente a control) (19). Los análisis de motivos se realizaron como en (1). En resumen, los fosfopéptidos investigados se agruparon primero según el motivo consenso/común que representan. A continuación, se calculó la media del  $\log_2$  del factor de proporción (frente a DMSO control) para cada grupo de motivos en cada condición.

Algoritmo de identificación de CTAM. Los sitios de fosforilación se filtraron primero para incluir solo aquellos que poseían una puntuación  $\delta$  de Mascot  $\geq 5$ . Los marcadores de actividad para cada inhibidor de quinasas se identificaron después seleccionando sitios de fosforilación con un  $\log_2$  de factor de proporción  $\leq 1$  (es decir, una reducción lineal del 50% en la abundancia) y  $P$  ajustado  $\leq 0,1$ , o con aquellos con un cambio menor ( $\log_2$  de factor de proporción  $\leq 0,75$ ), pero mayor significación ( $P$  ajustado  $\leq 0,05$ ) para cada inhibidor. Los sitios de fosforilación que cumplieron los umbrales anteriores en ambos tratamientos con inhibidores dirigidos a la misma quinasa se identificaron posteriormente como marcadores de actividad compuesto-diana (CTAM).

Ensayos de viabilidad celular. Las líneas celulares MCF7 precursoras, G1:3 y K1:3 se sembraron en placas de 96 pocillos a una densidad de aproximadamente 5.000 células.pocillo<sup>-1</sup>, por triplicado biológico. Tras 24 horas, las células se trataron después con las concentraciones/combinaciones indicadas de GDC-0941, KU-0063794, MK-2206, KN-93, PD-153035, GSK-1120212 o DMSO (como en la Tabla 1). Después de un tratamiento de 48 horas con los inhibidores, se determinó la proliferación celular con el ensayo MTS (Promega, WI, EE. UU.) o tinción con cristal violeta. Para los ensayos de MTS, se determinó la absorbancia (490 nm) de cada pocillo usando un espectrofotómetro después de una incubación de 120 minutos con el reactivo, adquiriéndose esta medición por duplicado. Para la tinción con cristal violeta, las células se fijaron con 100  $\mu$ l.pocillo<sup>-1</sup> de paraformaldehído (4 % p/v en PBS) durante 30 minutos en hielo, se tiñeron con 100  $\mu$ l.pocillo<sup>-1</sup> de tinción con cristal violeta (0,5 % p/v en MeOH al 20 %) durante 10 minutos a temperatura ambiente y se lavaron dos veces con ddH<sub>2</sub>O. Después de los lavados, el cristal violeta se resuspendió usando 100  $\mu$ l.pocillo<sup>-1</sup> de tampón de Sorensen (citrato de sodio 0,1 M, EtOH al 50 %, pH 4,2) y la absorbancia (540 nm) de cada pocillo se midió por duplicado. La relación de la absorbancia de cada punto de datos de concentración frente a las células tratadas con la absorbancia de control (DMSO) se calculó después, se transformó por log y se calculó la media de cada conjunto de repeticiones.

Tabla 1. Inhibidores de moléculas pequeñas utilizados en los experimentos

Inhibidor	Concentración usada ( $\mu$ M)	Diana primaria	Inespecíficos a partir de la base de datos de MRC (inhibición > del 50 % a la concentración dada)	ChEMBL ID	Inespecíficos para ChEMBL	Material de referencia
Inhibidor de Akt VIII	1	Akt1/2/3	CAMK1 (1 $\mu$ M)	258844	-	(20)
MK-2206	1	Akt1/2/3	-	1079175	-	(21,22)
KN-62	5	CAMK2 $\alpha$ / $\beta$ / $\gamma$ / $\delta$	PRAK, CAMK1, DYRK1A, Lck (10 $\mu$ M)	155333	P2RX7	(23)
KN-93	10	CAMK2 $\alpha$ / $\beta$ / $\gamma$ / $\delta$	TrkA, smMLCK (10 $\mu$ M)	28234	-	(24)
PD-168393	1	EGFR	-	285063	ERBB2/4, BMX, BTK, BLK, JAK3	(25)

ES 2 833 556 T3

(continuación)

Inhibidor	Concentración usada (µM)	Diana primaria	Inespecíficos a partir de la base de datos de MRC (inhibición > del 50 % a la concentración dada)	ChEMBL ID	Inespecíficos para ChEMBL	Material de referencia
PD-153035	10	EGFR	-	29197	ABL1, ERBB2, MKNK1, F16P1	(26)
inhibidor de ERK I	40	ERK1/2	-	1403932	-	(27)
inhibidor de ERK II (FR180204)	30	ERK1/2	-	259551	MAPK14	(28)
GSK-1120212	0,5	MEK1/2	-	-	-	(29, 30)
U0126	10	MEK1/2	MKK1 (10 µM)	100473	-	(30, 31)
KU-0063794	1	mTORC1/2	-	1078983	-	(32)
Torin-1	1	mTORC1/2	-	1256459	PI3Kα/δ /γ, p85α, PI3K-C2α/β, Vps34, DNA-PK	(33)
PF-4708671	20	p70S6K (S6K1)	MSK1, RSK1 (1 µM)	-	S6Kα5	(34)
DG2	5	p70S6K (S6K1)	-	1254209	-	(35)
GDC-0941	1	PI3K (clase I)	CLK2 (1 µM)	573393	Akt1, mTOR, HIPK2/3, MYLK4, CLK2, MAPK10, FLT3, JAK1, RIOK2	(36, 37)
PI-103	1	PI3K (clase I)	-	521851	mTOR, DNA-PK, Akt1	(38)
Gö-6976	1	PKCα/β/δ /γ	MKK1, BRSK2, ERK2/8, CAMKKα, MNK2, DYRK1A/3, SmMLCK, HIPK2, Aurora C, GSK3β, Aurora B, PIM1, MELK, AMPK, MARK3, CAMKKβ, S6K1, PDK1, CDK2, RSK1, PAK4/5/6, MSK1, PKD1, PIM3, RSK2, PRK2, PHK, CHK1, MST2 (0,1 µM)	302449	PKD1	(39)
Bisindolil maleimida -I	1	PKCα/θ/η /γ/ε	MKK1, ROCK2, Lck, PKA, Akt1, MAPKAP-K2, CHK1, PKD1, AMPK, S6K1, GSK3β, CDK2, SGK1, DYRK1A, MSK1, PHK, RSK1 (10 µM)	7463	CDK4/CCN D1, GSK3β, PIM1	(40)
Ácido ocadaico	1	PP2A	-	280487	-	(41)
Ácido cantárico	10	PP2A	-	275516	-	(42)
H-1152	5	ROCK	MNK1, MARK3, MELK, EPH-A2, RSK2, MSK1, BRSK2, FGF-R1, PKA, AMPK, RSK1, PHK, Aurora B/C, PRK2 (1 µM)	406821	PKA	(43)

(continuación)

Inhibidor	Concentración usada ( $\mu\text{M}$ )	Diana primaria	Inespecíficos a partir de la base de datos de MRC (inhibición > del 50 % a la concentración dada)	ChEMBL ID	Inespecíficos para ChEMBL	Material de referencia
Y-27632	10	ROCK	PHK, MST2, RSK2, MSK1, MNK1, RSK1, AMPK, PRK2 (1 $\mu\text{M}$ )	-	PKC $\epsilon$	(44)

Ensayos de proliferación celular. Las líneas celulares MCF7 precursoras, G1:3 y K1:3 se sembraron en placas de 12 pocillos a una densidad de aproximadamente 80.000 células.pocillo<sup>-1</sup>, por triplicado biológico. Después del período de tiempo indicado, las células se lavaron con PBS, se tripsinizaron y se contaron usando un contador celular Beckman Coulter Vi-CELL XR. Los polinomios de segundo grado se ajustaron a los datos utilizando el paquete *ggplot2* (v0.9.3.1) dentro de R (v3.0.0).

Inmunotransferencia. Las células se lisaron en tampón de lisis (Tris-HCl 50 mM [pH 7,4], EDTA 1 mM, NaCl 150 mM, Triton-X100 al 1 %; suplementado con Na<sub>3</sub>O<sub>4</sub> 1 mM, NaF 1 mM,  $\beta$ -glicerol fosfato 1 mM, Na<sub>2</sub>P<sub>2</sub>H<sub>2</sub>O<sub>7</sub> 2,5 mM, PMSF 1 mM y cóctel de inhibidores de proteasas 1X [Sigma Aldrich]). Las muestras se resolvieron mediante electroforesis en gel de SDS-poliacrilamida utilizando geles precolados con gradiente de 12 % o 4-15 % (BioRad). Las proteínas se transfirieron a membranas de PVDF utilizando el sistema *BioRad* Trans-Blot Turbo según las instrucciones del fabricante. Una vez transferidas, se bloquearon las membranas con leche desnatada en polvo al 5 % p/v en solución salina tamponada con Tris suplementada con Tween-20 al 0,1 %. Las membranas bloqueadas se incubaron con anticuerpos primarios y secundarios y se desarrollaron con sustrato quimioluminiscente SuperSignal West Pico (Thermo Scientific). Los anticuerpos primarios se usaron a dilución 1:500 o 1:1000 y los anticuerpos secundarios se usaron a dilución 1:5000.

Referencias complementarias (se aplican solo a los Materiales y Métodos Detallados, anteriores)

- Casado P, *et al.* (2013) Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* 6(268):rs6.
- Montoya A, Beltran L, Casado P, Rodriguez-Prados JC, & Cutillas PR (2011) Characterization of a TiO(2) enrichment method for label-free quantitative phosphoproteomics. *Methods* 54(4):370-378.
- Casado P & Cutillas PR (2011) A self-validating quantitative mass spectrometry method for assessing the accuracy of high-content phosphoproteomic experiments. *Mol. Cell. Proteomics* 10(1):M110 003079.
- Cutillas PR & Vanhaesebroeck B (2007) Quantitative profile of five murine core proteomes using label-free functional proteomics. *Mol. Cell. Proteomics* 6(9):1560-1573.
- Tsou CC, *et al.* (2010) IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol. Cell. Proteomics* 9(1):131-144.
- Mann B, *et al.* (2008) ProteinQuant Suite: a bundle of automated software tools for label-free quantitative proteomics. *Rapid Commun. Mass Spectrom.* 22(23):3823-3834.
- Escher C, *et al.* (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 12(8): 1111-1121.
- Bolstad BM, Irizarry RA, Astrand M, & Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185-193.
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Molec. Biol.* 3(1):Artículo3.
- Smyth GK (2005) *Limma: linear models for microarray data* (Springer New York).
- R Core Team (2013) *R: A language and environment for statistical computing* (R Foundation for Statistical Computing).
- Wickham H (2009) *ggplot2: elegant graphics for data analysis* (Springer New York).
- Warnes GR *et al.* (2013) *gplots: Various R programming tools for plotting data.*
- Wickham H (2007) Reshaping data with the reshape package. *J. Stat. Softw.* 21(12):1-20.
- Csardi G & Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*: 1695.
- Shannon P, *et al.* (2013) RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics* 14:217.
- Shannon P, *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11):2498-2504.
- Gobbi A, *et al.* (2014) Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics* 30(17):i617-623.
- Kim SY & Volsky DJ (2005) PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics* 6:144.
- Lindsley CW, *et al.* (2005) Allosteric Akt (PKB) inhibitors: discovery and SAR of isozyme selective inhibitors. *Bioorg. Med. Chem. Lett.* 15(3):761-764.
- Rehan M, Beg MA, Parveen S, Damanhoury GA, & Zaher GF (2014) Computational insights into the inhibitory mechanism of human AKT1 by an orally active inhibitor, MK-2206. *PLoS ONE* 9(10):e109705.
- Hirai H, *et al.* (2010) MK-2206, an allosteric Akt inhibitor, enhances the antitumor efficacy by standard

- chemotherapeutic agents or molecular targeted drugs *in vitro* and *in vivo*. Mol. Cancer Ther. 9(7):1956-1967.
23. Sumi M., *et al.* (1991) The newly synthesized selective Ca<sup>2+</sup>/calmodulin dependent protein kinase-II inhibitor KN-93 reduces dopamine contents in PC12H cells. Biochem. Biophys. Res. Commun. 181(3):968-975.
24. Tokumitsu H, *et al.* (1990) KN-62, 1-N,O-bis(5-isoquinolinesulfonyl)-N-methyl-L-tyrosyl-4-phenylpiperazin E, a specific inhibitor of Ca<sup>2+</sup>/calmodulin-dependent protein kinase-II. J. Biol. Chem. 265(8):4315-4320.
- 5 25. Fry DW, *et al.* (1998) Specific, irreversible inactivation of the epidermal growth factor receptor and erbB2, by a new class of tyrosine kinase inhibitor. Proc. Natl. Acad. Sci. U.S.A. 95(20):12022-12027.
26. Bos M, *et al.* (1997) PD153035, a tyrosine kinase inhibitor, prevents epidermal growth factor receptor activation and inhibits growth of cancer cells in a receptor number-dependent manner. Clin. Cancer Res. 3(11):2099-2106.
- 10 27. Hancock CN, *et al.* (2005) Identification of novel extracellular signal-regulated kinase docking domain inhibitors. J. Med. Chem. 48(14):4586-4595.
28. Ohori M, *et al.* (2005) Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex. Biochem. Biophys. Res. Commun. 336(1):357-363.
29. Gilmartin AG, *et al.* (2011) GSK1120212 (JTP-74057) Is an Inhibitor of MEK Activity and Activation with Favorable Pharmacokinetic Properties for Sustained *In Vivo* Pathway Inhibition. Clin. Cancer Res. 17(5):989-1000.
- 15 30. Roberts PJ & Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 26(22):3291-3310.
31. Favata MF, *et al.* (1998) Identification of a novel inhibitor of mitogen-activated protein kinase. J. Biol. Chem. 273(29):18623-18632 (1998).
- 20 32. Garcia-Martinez JM, *et al.* (2009) Ku-0063794 is a specific inhibitor of the mammalian target of rapamycin (mTOR). Biochem. J. 421(1):29-42.
33. Liu QS, *et al.* (2010) Discovery of 1-(4-(4-Propionylpiperazin-1-yl)-3-(trifluoromethyl)phenyl)-9-(quinolin-3-yl)benzo h 1,6 naphthyridin-2(1H)-one as a Highly Potent, Selective Mammalian Target of Rapamycin (mTOR) Inhibitor for the Treatment of Cancer. J. Med. Chem. 53(19):7146-7155.
- 25 34. Pearce LR, *et al.* (2010) Characterization of PF-4708671, a novel and highly specific inhibitor of p70 ribosomal S6 kinase (S6K1). Biochem. J. 431(2):245-255.
35. Okuzumi T, *et al.* (2009) Inhibitor hijacking of Akt activation. Nat. Chem. Biol. 5(7):484-493.
36. Raynaud FI, *et al.* (2009) Biological properties of potent inhibitors of class I phosphatidylinositide 3-kinases: from PI-103 through PI-540, PI-620 to the oral agent GDC-0941. Mol. Cancer Ther. 8(7):1725-1738.
- 30 37. Miller TW, Rexer BN, Garrett JT, & Arteaga CL (2011) Mutations in the phosphatidylinositol 3-kinase pathway: role in tumor progression and therapeutic implications in breast cancer. Breast Cancer Res. 13(6):224.
38. Raynaud FI, *et al.* (2007) Pharmacologic characterization of a potent inhibitor of class I phosphatidylinositide 3-kinases. Cancer Res. 67(12):5840-5850.
- 35 39. Martinybaron G, *et al.* (1993) Selective inhibition of protein kinase C isozymes by the indolocarbazole Gö 6976. J. Biol. Chem. 268(13):9194-9197.
40. Toullec D, *et al.* (1991) The bisindolylmaleimide GF-109203X is a potent and selective inhibitor of protein kinase C. J. Biol. Chem. 266(24):15771-15781.
41. McCluskey A, Sim ATR, & Sakoff JA (2002) Serine-threonine protein phosphatase inhibitors: Development of potential therapeutic strategies. J. Med. Chem. 45(6):1151-1175.
- 40 42. Li YM & Casida JE (1992) Cantharidin-binding protein - identification as protein phosphatase-2A. Proc. Natl. Acad. Sci. U.S.A. 89(24):11867-11870.
43. Tamura M, *et al.* (2005) Development of specific Rho-kinase inhibitors and their clinical application. Biochim. Biophys. Acta-Proteins and Proteomics 1754(1-2):245-252.
44. Ishizaki T, *et al.* (2000) Pharmacological properties of Y-27632, a specific inhibitor of Rho-associated kinases. Mol. Pharmacol. 57(5):976-983.
- 45

## Resultados

Los presentes inventores se propusieron clasificar los sitios de fosforilación en grupos definidos por sus patrones de modulación en respuesta a los inhibidores de la señalización celular. Se trataron las células MCF7 con inhibidores de moléculas pequeñas individuales contra un conjunto de quinasas y se midieron los cambios resultantes en la fosforilación utilizando fosfoproteómica basada en EM (Fig. 1A). Las condiciones utilizadas para el experimento fueron veinte inhibidores de quinasas estructuralmente distintos, dos inhibidores de fosfatasas o vehículo control de DMSO (Fig. 1B). Las quinasas dirigidas en los experimentos, junto con los inhibidores utilizados, se eligieron en función de su participación conocida en el factor de crecimiento y la señalización metabólica, y su potencial terapéutico actual.

50

55

Los experimentos de EM dieron como resultado la identificación de un total de 13.405 iones de fosfopéptidos únicos en las seis repeticiones analíticas por condición (tres técnicas y dos biológicas). En la figura 5 se muestra un resumen de control de calidad de estos datos. Cada fosfopéptido se cuantificó en todas las condiciones experimentales usando una metodología sin marcadores descrita previamente (18, 19), generando 1.930.320 puntos de datos.

60

Los datos de fosfoproteómica permiten la clasificación de los inhibidores de quinasas en función de las dianas que inhiben. Se observaron 4.651 sitios de fosforilación significativamente reducidos en abundancia por al menos un inhibidor de quinasas ( $P$  ajustado  $\leq 0,05$ ). Para evaluar los efectos globales de los inhibidores sobre estos sitios, se utilizó el análisis de componentes principales (PCA). Este método de análisis estadístico multivariante permite la separación de condiciones experimentales basándose en la estructura general de los datos subyacentes. El PCA de

65

los fosfoproteomas tratados con inhibidor demostró que los inhibidores dirigidos contra la misma quinasa estaban más cerca uno del otro en el espacio del componente principal que el resto de los inhibidores (Fig. 1C), lo que indica que los inhibidores contra la misma quinasa produjeron efectos similares sobre la fosforilación global. La única excepción a esta observación fueron los inhibidores de ERK; estos estaban cerca en la dimensión PC1 pero no en la PC2, lo que sugiere que estos tenían efectos cuantitativos ligeramente diferentes sobre el fosfoproteoma. Mientras que los inhibidores contra las quinasas relacionadas con la cascada de señalización de MAPK (EGFR, MEK y ERK) se separaron de los relacionados con el eje de señalización PI3K/mTOR (PI3K, mTOR, p70S6K), los inhibidores de Akt se asociaron más estrechamente con los inhibidores de EGFR y CAMK2 que los inhibidores de su conocido activador cadena arriba, PI3K. Como cabría esperar, los inhibidores dirigidos a PP2A (una proteína fosfatasa) se separaron bien de los inhibidores de quinasas. El análisis utilizando una matriz de correlación reforzó las relaciones observadas entre los inhibidores de mTOR, Akt y PI3K y entre los pares de inhibidores EGRF, ERK y MEK (Fig. 1D). La significación estadística de cada correlación se muestra en la Fig. 6. El análisis de motivos (14) reveló además que los pares de inhibidores mostraron impactos fuertemente correlacionados sobre los motivos de fosforilación específicos mientras reflejaban las relaciones observadas en las Fig.1C y 1D (figura 1E). En conjunto, estos datos muestran que los inhibidores contra las mismas quinasas produjeron cambios más similares en los fosfoproteomas que en el resto de los inhibidores y, con la excepción de Akt, aquellos contra las mismas vías canónicas también afectaron un conjunto común de sitios de fosforilación.

Deducir marcadores de actividad de dianas de inhibidores a partir de datos de fosfoproteómica. Para proporcionar un nivel adicional de clasificación al conjunto de datos, se agruparon aún más los sitios de fosforilación en función de su comportamiento en tratamiento con inhibidores contra la misma quinasa. El número de sitios de fosforilación seleccionados en esta etapa dependía de la rigurosidad de los umbrales utilizados para la selección (Fig. 7). Sin embargo, lo más importante es que se seleccionaron solo aquellos sitios de fosforilación que alcanzaron los umbrales estadísticos requeridos en ambos tratamientos con inhibidores dirigidos a la misma quinasa. La selección de sitios de fosforilación inhibidos por compuestos estructuralmente distintos dirigidos a la misma quinasa debería dar como resultado conjuntos de datos enriquecidos en sitios de fosforilación específicos de la quinasa deseada. Esto se ilustra en la Fig. 2A para los sitios de fosforilación modulados por los dos inhibidores de Akt diferentes (MK-2206 e inhibidor de Akt VIII), que muestra los sitios de fosforilación inhibidos por ambos inhibidores (puntos de datos sombreados oscuro debajo de la línea punteada horizontal inferior y a la izquierda de la línea punteada más a la izquierda ( $FR \leq -1,0$ ) en la Fig. 2A), así como los sitios específicamente inhibidos por un compuesto pero no por el otro (puntos de datos sombreados oscuro por encima de la línea punteada horizontal inferior y debajo de la siguiente línea horizontal punteada más alta; puntos de datos sombreados oscuro a la derecha de la línea vertical punteada más a la izquierda y a la izquierda de la segunda línea vertical punteada a la izquierda ( $-0,75 < FR < -1,0$ ) en la figura 2A). Se ha planteado la hipótesis de que los sitios inhibidos por el inhibidor de Akt MK-2206 pero no por el inhibidor de Akt VIII, y *viceversa*, eran efectos inespecíficos, mientras que los inhibidos por ambos compuestos tenían más probabilidades de estar realmente cadena abajo de Akt. Este análisis se realizó para cada una de las 10 quinasas dirigidas en el estudio (Fig. 1A) y reveló 610 sitios de fosforilación reducidos en abundancia por al menos un par de inhibidores (es decir, por ambos inhibidores contra la misma quinasa). Estos sitios, aunque no necesariamente fosforilados directamente por la quinasa diana deseada, ya que podrían ser fosforilados por quinasas que actúan cadena abajo o por quinasas estrechamente relacionadas, son lecturas de las quinasas reales afectadas por el inhibidor/compuesto, y por eso los presentes inventores nos referimos a ellas como marcadores de actividad compuesto-diana (CTAM).

Inferir topología de red de señalización a partir de datos de fosfoproteómica. La visualización de los 610 sitios de fosforilación de CTAM identificados reveló simultáneamente que un gran número de ellos se identificaron como marcadores de más de un par compuesto-diana (Fig. 2B). Por lo tanto, para investigar adicionalmente las relaciones entre los pares de inhibidores, y para permitir la inferencia de la topología de la red de señalización a partir de los datos, los sitios de fosforilación de los 610 CTAM se clasificaron adicionalmente basándose en si estos fueron inhibidos por uno o más pares de inhibidores. De este análisis surgieron varios patrones conocidos de topología de señalización de quinasas (Fig. 2C). Por ejemplo, se identificaron 41 sitios de fosforilación que fueron inhibidos por los pares de inhibidores contra Akt, mTOR, p70S6K y PI3K (Fig. 2C). De acuerdo con el conocimiento previo, estos sitios incluyeron aquellos en BAD, Ser<sup>427</sup> en KS6B1 (p70S6K1) y Thr<sup>1135</sup> en RICTOR (Fig. 2C) (20-22). Junto a estos, los sitios que aún no se han anotado funcionalmente también estuvieron presentes en este grupo, y juntos estos 41 sitios se clasificaron como miembros del grupo de CTAM Akt-mTOR-PI3K-p70S6K. Análogamente, los sitios modulados por los pares de inhibidores de Akt, mTOR y PI3K, pero no por el par de inhibidores de p70S6K, incluyeron GSK3 $\beta$  en Ser<sup>9</sup>, Myc en Ser<sup>62</sup> y AKTS1 (también conocido como PRAS40) en Ser<sup>183</sup> (Fig. 8); se encontró que un total de 55 sitios de fosforilación tenían este patrón de inhibición y definieron un grupo Akt-mTOR-PI3K que es independiente de p70S6K. Además de estas conocidas cascadas de quinasas, también encontramos evidencia de la existencia de relaciones aún no caracterizadas entre las quinasas dirigidas por los inhibidores y/o los inhibidores mismos; los ejemplos incluyen sitios modulados por pares de inhibidores de mTOR y PI3K sin la participación de Akt o p70S6K (37 sustratos, Fig. 2B-C). Como ilustra la Fig. 2B-C, también encontramos evidencia de sitios inhibidos por ambos inhibidores de Akt pero no afectados por PI3K y otros inhibidores (284 sustratos) y sitios de PI3K independientes de Akt y mTOR (33 sustratos). En general, los 610 marcadores de actividad de los sitios de fosforilación encontrados en este estudio (Tabla 1) se agruparon en 55 grupos de CTAM.

Visualizar los datos como un grafo de red bipartito, no dirigido (Fig. 2D) reveló adicionalmente la forma en que los inhibidores de quinasas investigados se relacionaban entre sí en la red de señalización. Como se esperaba y en

concordancia con los datos publicados y sus asociaciones canónicas, los pares de inhibidores de Akt, PI3K, p70S6K y mTOR afectaron un gran número de sitios de fosforilación comunes y, por lo tanto, se agruparon. Análogamente, los pares de inhibidores dirigidos a la ruta de MAPK (EGFR, MEK y ERK) también se agruparon y con CAMK2 (Fig. 2D). La aleatorización de la topología de la red reveló que no era probable que estas asociaciones ocurrieran por casualidad (Fig. 9). Por lo tanto, estos datos muestran que, aunque las asociaciones canónicas entre las quinasas estaban bien representadas en el conjunto de datos de los presentes inventores (por ejemplo, EGFR-MEK-ERK y PI3K-Akt-mTOR-p70S6K), la existencia de rutas de señalización inesperadas también surgió de estos datos, incluida la existencia de señalización PI3K-mTOR independiente de Akt. Por otra parte, este análisis demostró el alto grado de conectividad entre estas quinasas.

Caracterización del comportamiento de los grupos de CTAM identificados y sondeo de la plasticidad de la red. A continuación se buscó confirmar si los grupos de CTAM podrían usarse para medir la activación bioquímica de las vías dentro de la red, y así proporcionar una instantánea del estado de activación de la red en cualquier momento, en cualquier condición. Se ha planteado la hipótesis de que, si estos grupos proporcionan lecturas confiables de la actividad de la rama de la red, cada grupo de CTAM debe mostrar los cambios de comportamiento esperados cuando la red se perturba o estimula en condiciones experimentales bien caracterizadas. Además, se razonó que los miembros individuales dentro de cada grupo deberían demostrar un comportamiento cuantitativo similar entre sí. Por lo tanto, se monitorizó la dinámica de la fosforilación de los grupos de CTAM a través de las células tratadas con EGF o IGF-1 en cinco puntos de tiempo independientes. En la figura 10 se muestra un resumen de control de calidad para este conjunto de datos.

Se observó que las dinámicas de fosforilación temporal de los grupos de CTAM comúnmente asociados con la señalización de EGFR e IGF-1 R (en relación con el control de 0 minutos en cada caso) estaba en línea con los efectos indicados previamente de EGF e IGF-1 sobre la señalización de quinasas (Fig. 3A) (23, 24). Por ejemplo, de acuerdo con la dinámica temporal conocida de la activación de la ruta MAPK, el grupo EGFR-MEK se sometió a una regulación ascendente aguda significativa tras la estimulación con ambos factores de crecimiento durante 5 minutos, antes de comenzar a disminuir a un nivel inferior a los 60 minutos (Fig. 3A-B). Estos datos estaban de acuerdo con los datos individuales de EM y transferencia western para los sitios canónicos de MAPK sensibles a EGFR (Thr<sup>202</sup>/Tyr<sup>204</sup>) y Akt (Ser<sup>473</sup>) (Fig. 11). La mediana de las desviaciones estándar relativas (es decir, el coeficiente de variación) de los sitios de fosforilación individuales dentro de los grupos de CTAM fueron 0,454 y 0,518 para EGF e IGF-1 respectivamente (Fig. 3C), lo que refleja que estos se comportaron de manera similar en la estimulación celular con los dos factores de crecimiento. En conjunto, los datos mostrados en la Fig. 3 proporcionan evidencia para apoyar la noción de que los grupos de CTAM eran lecturas de la activación funcional de las ramas dentro de la red.

Análisis de plasticidad de la red en modelos de resistencia adquirida a inhibidores de quinasas. Para investigar adicionalmente la plasticidad de la señalización de quinasas en la red de señalización de los presentes inventores definida por CTAM, se midieron los sitios de fosforilación que definen la red en modelos de líneas celulares de cáncer de resistencia adquirida a dos inhibidores de quinasas en el desarrollo clínico; en concreto, GDC-0941 (un inhibidor pan de PI3K de clase I) y KU-0063794 (un inhibidor de mTORC1/2) (25, 26). Se obtuvieron seis cultivos celulares independientes resistentes a cada uno de los inhibidores en comparación con las células precursoras de las que derivaron (tres por fármaco: MCF7-G y MCF7-K resistentes a GDC-041 y KU-0063794, respectivamente). Para conseguir esto, las células se expusieron de forma crónica a una concentración creciente del inhibidor relevante hasta un máximo de 1  $\mu$ M. Las células se expusieron inicialmente con una baja concentración de cada fármaco (100 nM) para no sesgar la selección por resistencia para las células intrínsecamente resistentes. Las líneas celulares resultantes pudieron proliferar en presencia de 1  $\mu$ M de inhibidor, mientras que las células precursoras no pudieron hacerlo en las mismas condiciones (Fig. 4A-B). Se cuantificaron los fosfoproteomas de estas células y se normalizaron estas mediciones a proteína total analizando simultáneamente el proteoma total. En la Fig. 12 se muestra un resumen de los datos cuantitativos y cualitativos.

El análisis de la red de quinasas en presencia de inhibidores reveló que la gran mayoría de los grupos de CTAM que involucraban a mTOR estaban regulados negativamente en todas las líneas celulares resistentes a inhibidores de mTOR (MCF7-K) (flechas verde y naranja en la Fig. 4C). Análogamente, los grupos de CTAM que contenían PI3K estaban regulados negativamente en todas las líneas celulares resistentes a inhibidores de PI3K (MCF7-G) (flechas verdes en la Fig. 4C). Se confirmaron estos datos midiendo marcadores bien conocidos de las actividades de las rutas, que mostraron que los resultados de los presentes inventores fueron consistentes con los niveles de los sitios clave de fosforilación reguladora que rigen estas rutas en Akt (Ser<sup>473</sup>) y p70S6K (Thr<sup>389</sup>), como se determina por transferencia Western (Fig. 13). Estos datos sugirieron que las vías dirigidas por los inhibidores permanecieron inhibidas en células resistentes en presencia del fármaco. Por lo tanto, se razonó que la resistencia no era el resultado de diferencias en la forma en que las células resistentes metabolizaron los inhibidores y que, en cambio, de acuerdo con otros estudios (27), la resistencia probablemente surgió como consecuencia de una nueva reconexión de la señalización de las quinasas. De interés, esta nueva reconexión fue notablemente diferente entre las líneas celulares MCF7-K y MCF7-G (Fig. 4C), lo que sugiere que los mecanismos de resistencia que habían evolucionado contra el inhibidor de mTOR eran distintos a los desarrollados en respuesta al inhibidor de PI3K. Esta hipótesis se reforzó mediante el uso de un análisis multivariado imparcial de los datos de fosfoproteómica normalizados, que destacó las diferencias entre las células precursoras y resistentes, y las diferencias entre la reconexión de las células MCF7-K y MCF7-G, ya que estas se separaron claramente en el espacio del componente principal (Fig. 4D). Inesperadamente, sin embargo, las células

resistentes al mismo inhibidor también se separaron en el espacio del PC (por sus siglas en inglés, *principal component*), lo que sugiere que el estado de activación de la red de señalización era heterogéneo entre líneas resistentes individuales (Fig. 4C-D) a pesar de que derivan del mismo cultivo precursor, en el mismo número de pase, y estando expuestas a condiciones experimentales idénticas durante la misma cantidad de tiempo.

Debido a que las líneas celulares resistentes parecían diferir en la forma en que habían reconectado su red de señalización en comparación con las células precursoras y entre sí (Fig. 4C-D), se planteó la hipótesis de que cada línea celular debería responder de manera diferente a un grupo de inhibidores de quinasas de moléculas pequeñas solos y en combinación, ya que su respuesta estaría en función de la actividad de su red de señalización. Para probar esta predicción, se trató cada una de las líneas celulares con un conjunto de inhibidores de moléculas pequeñas (dirigidos a PI3K, mTOR, CAMK2, Akt, MEK y EGFR) y se midió su viabilidad celular relativa utilizando los ensayos MTS y con cristal violeta (Fig. 14A y B, respectivamente). Se eligieron estos inhibidores porque los CTAM de sus quinasas asociadas aumentaron en abundancia en algunas de las células resistentes con respecto a las precursoras (Fig. 4C), lo que sugiere que estas quinasas pueden estar implicadas en el fenotipo de resistencia. Un análisis imparcial y multivariado de los datos resultantes reveló que las células resistentes y precursoras respondieron de manera diferente a los inhibidores, ya que se separaron en el espacio del PC. Por otra parte, este análisis separó las células MCF7-G y -K y las líneas celulares resistentes individuales entre sí de una manera que refleja lo observado en el PCA de los datos de fosfoproteómica (Fig. 4E-F). En conjunto, estos datos indican que la reconexión heterogénea de la red de señalización en las células resistentes observada por el análisis de CTAM (Fig. 4C) dio como resultado diferencias funcionales en la forma en que las células respondieron a las perturbaciones de la red (Fig. 4E-F).

#### Análisis

En este estudio, primero se realizó un análisis exhaustivo de las conexiones que existen entre los nodos de la red de quinasas PI3K-MEK (Figs. 1 y 2). Este estudio inicial reveló los vínculos esperados e inesperados entre las quinasas, las vías de señalización y los agentes farmacológicos que las dirigen. Por ejemplo, las conocidas relaciones PI3K-Akt-mTOR-p70S6K, PI3K-Akt-mTOR y MEK-ERK estaban bien representadas en nuestros datos; sin embargo, también se encontró evidencia de conexiones aún no caracterizadas entre dianas de inhibidores de quinasas, tal como las definidas por asociaciones mTOR-PI3K sin la participación de Akt y sitios dependientes del inhibidor de Akt pero independientes del inhibidor de PI3K. En general, los datos de los presentes inventores ejemplifican la compleja relación entre las quinasas en las redes de señalización e ilustran que el conocimiento de los presentes inventores de esta complejidad aún es muy limitado.

Una ventaja de definir rutas de señalización utilizando un conjunto de datos experimentales derivados de un sistema definido, en contraste con los enfoques que recopilan información de la literatura (28, 29), es que los eventos de señalización celular, con frecuencia, dependen del tipo de célula y del contexto (30). Por lo tanto, las redes de señalización "promediadas", derivadas de tipos de células y organismos dispares, tal como se muestra en los esquemas de la vía de señalización canónica, no siempre son representativos de cómo las redes de señalización están conectadas de hecho en sistemas celulares específicos. Si bien se han realizado esfuerzos para superponer datos transcripcionales empíricos en estas redes promediadas (31), un aspecto clave del estudio de los presentes inventores es que no solo brindamos evidencia de rutas de señalización aún no caracterizadas sino que también identificamos marcadores de sitios de fosforilación de actividades específicas de dichas rutas para el modelo de línea celular de los presentes inventores, que después podría usarse para medir las dinámicas y los circuitos de la red de quinasas de manera sistemática. El enfoque de CTAM para definir las ramas de la red de señalización, que después se puede medir en experimentos posteriores, tiene similitudes conceptuales con los enfoques que derivan la información biológica celular de los patrones de expresión génica al examinar cómo estos patrones se correlacionan con los compendios de perfiles obtenidos de los experimentos sistemáticos de inactivación génica (32). La observación de que los grupos de CTAM se modularon por factores de crecimiento con la cinética esperada [(23, 24, Fig. 3) y que estos cambios fueron similares para los miembros de dichos grupos (Fig. 3) proporcionó evidencia que sugiere que estos grupos de CTAM son lecturas bioquímicas de actividad de señalización.

Las redes de señalización no son estáticas, sino estructuras altamente dinámicas que son extremadamente plásticas en respuesta a estímulos externos. Una comparación de la red entre las células precursoras y las células resistentes a un inhibidor de PI3K o mTORC1/2 reveló diferencias generalizadas en la abundancia del grupo de CTAM en tres cultivos de células resistentes por inhibidor (Fig. 4C). Por lo tanto, los datos de los presentes inventores son consistentes con los estudios publicados que muestran que la señalización de las quinasas se remodela en respuesta a la inhibición crónica de las quinasas (33, 34). Debido a la profundidad de los análisis de los presentes inventores, sin embargo, los datos de los presentes inventores enfatizan la medida en que las redes de señalización se modulan en su conjunto para superar la inhibición crónica de los nodos individuales. Simultáneamente con estas observaciones, la comparación de los datos sobre el estado de la red (fosfoproteómica) y los fenotipos (respuestas a los inhibidores) para cada línea celular utilizando análisis multivariados indicaron que la actividad de la red de señalización era un reflejo de los fenotipos recién adquiridos de las células. Esto se evidenció en los PCA de todos estos conjuntos dispares de datos, que separaron las células resistentes de las precursoras, y las células MCF7-K de las células MCF7-G de manera similar (Fig. 4D-F). Por lo tanto, los datos de los presentes inventores acentúan la compleja relación entre PI3K y mTORC1/2, ya que las células tratadas con los dos inhibidores cambiaron su señalización de manera diferente, lo que sugiere diferentes mecanismos de resistencia en respuesta a los dos inhibidores (Fig. 4D-F). Esto fue

inesperado ya que PI3K y mTOR con frecuencia se colocan en la misma vía de señalización canónica. Sin embargo, estos datos son consistentes con la observación de que los sitios de fosforilación modulados por los inhibidores de mTORC1/2 o de PI3K solo se solapan parcialmente (Fig. 2E), y con un estudio reciente que destaca la independencia de mTOR de PI3K en algunos sistemas (35). El simple mantenimiento de las células en cultivo podría haber contribuido a la evolución de las redes celulares que se muestran en la Fig. 4; sin embargo, la observación de que la exposición crónica a los inhibidores de mTORC1/2 o PI3K produjo marcadas diferencias en la remodelación de la red argumenta que la exposición a los fármacos tuvo un efecto mayor sobre cómo evolucionó la señalización que lo que se observaría como resultado del cultivo a largo plazo sin la aplicación de una presión tan selectiva.

Publicaciones recientes indican que, en lugar de que exista un mecanismo único de resistencia adquirida para una terapia dada, hay una gran cantidad de formas en que las redes de señalización pueden volver a conectarse en las células cancerosas a medida que se vuelven resistentes a las terapias dirigidas (36, 37). Sin embargo, no se sabe si la forma en que las células cancerosas desarrollan resistencia está predeterminada por la impronta molecular de las células cancerosas en la presentación (es decir, antes de que se produzca el fenotipo de resistencia). Esta comprensión tiene implicaciones potenciales para la terapéutica del cáncer, ya que si las vías evolutivas que conducen a la adquisición de resistencia dependieran de factores genómicos u otros factores moleculares presentes en las células tumorales en la presentación, entonces, el análisis de las condiciones iniciales de las células cancerosas podría, al menos en principio, utilizarse para predecir el mecanismo de resistencia que con mayor probabilidad se producirá en dicho tumor.

Contrariamente a esta idea, los datos de los presentes inventores sugieren que puede no ser posible predecir el mecanismo particular de resistencia de un tumor dado analizándolo a nivel molecular en la presentación. De hecho, en el estudio de los presentes inventores, la misma línea celular se dividió en seis poblaciones idénticas, que después se mantuvieron en las mismas concentraciones de inhibidores por el mismo operador durante la misma cantidad de tiempo y número de pasadas. Inesperadamente, los perfiles de la red de CTAM fueron altamente heterogéneos a través de células resistentes al mismo compuesto (Fig. 4C-D). Esta heterogeneidad en el estado de la red tuvo una consecuencia funcional en que las células resistentes al mismo compuesto respondieron de manera diferente a los inhibidores de la señalización al nivel de proliferación celular (Fig. 4E-F). La heterogeneidad intratumoral y la divergencia evolutiva de los sistemas inicialmente clónicos se ha documentado a nivel genético en los sistemas de mamíferos y bacterias (38, 39). Aunque no se puede excluir por completo el impacto del cultivo celular a largo plazo, el estudio de los presentes inventores sugiere que el tratamiento crónico con inhibidores de quinasas dirigidos influye profundamente en la divergencia de la señalización de la red de quinasas. Por lo tanto, como las condiciones experimentales idénticas pueden dar como resultado la evolución de distintas redes de señalización (Fig. 4C-D) y fenotipos de resistencia a los fármacos (Fig. 4E-F), quizás debido a los efectos estocásticos, las condiciones iniciales del sistema pueden no ser predictores precisos de la ruta evolutiva que puede conducir a la resistencia. Por lo tanto, si la adquisición de resistencia es realmente indeterminista y no puede predecirse al inicio de la enfermedad, la identificación de los mecanismos de resistencia irrelevantes para cada paciente individual probablemente requerirá la capacidad de medir la red de señalización en tumores individuales después de que haya surgido la resistencia con profundidad y sin preconcepción de cómo la señalización puede haberse reconectado como resultado de la terapia.

Referencias (se aplican al Ejemplo excepto para la sección Materiales y Métodos Detallados, anterior)

1. Du W & Elemento O (2014) Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* [Epub ahead of print].
2. Papin JA, Hunter T, Palsson BO, & Subramaniam S (2005) Reconstruction of cellular signaling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* 6(2):99-111.
3. Bodenmiller B, *et al.* (2010) Phosphoproteomic Analysis Reveals Interconnected System-Wide Responses to Perturbations of Kinases and Phosphatases in Yeast. *Sci. Signal* 3(153):rs4.
4. Jorgensen C & Linding R (2010) Simplistic pathways or complex networks? *Curr. Opin. Genet. Dev.* 20(1):15-22.
5. Hsu PP, *et al.* (2011) The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* 332(6035):1317-1322.
6. Goltsov A, *et al.* (2012) Features of the reversible sensitivity-resistance transition in PI3K/PTEN/AKT signaling network after HER2 inhibition. *Cell. Signal.* 24(2):493-504.
7. Goltsov A, *et al.* (2011) Compensatory effects in the PI3K/PTEN/AKT signaling network following receptor tyrosine kinase inhibition. *Cell. Signal.* 23(2):407-416.
8. Kholodenko BN, *et al.* (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U.S.A.* 99(20):12841-12846.
9. Mukherjee S & Speed TP (2008) Network inference using informative priors. *Proc. Natl. Acad. Sci. U.S.A.* 105(38):14313-14318.
10. Prill RJ, Saez-Rodríguez J, Alexopoulos LG, Sorger PK, & Stolovitzky G (2011) Crowdsourcing network inference: the DREAM predictive signaling network challenge. *Sci. Signal* 4(189):mr7.
11. Linding R, *et al.* (2007) Systematic Discovery of *In Vivo* Phosphorylation Networks. *Cell* 129(7):1415-1426.
12. Carlson SM, *et al.* (2011) Large-Scale Discovery of ERK2 Substrates Identifies ERK-Mediated Transcriptional Regulation by ETV3. *Sci. Signal.* 4(196).
13. Bensimon A, Heck AJ, & Aebersold R (2012) Mass spectrometry-based proteomics and network biology. *Annu.*

Rev. Biochem. 81:379-405.

14. Casado P, *et al.* (2013) Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* 6(268):rs6.
15. Posch C, *et al.* (2013) Combined targeting of MEK and PI3K/mTOR effector pathways is necessary to effectively inhibit NRAS mutant melanoma *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 110(10):4015-4020.
- 5 16. Renshaw J, *et al.* (2013) Dual blockade of the PI3K/AKT/mTOR (AZD8055) and RAS/MEK/ERK (AZD6244) pathways synergistically inhibits rhabdomyosarcoma cell growth *in vitro* and *in vivo*. *Clin. Cancer Res.* 19(21):5940-5951.
17. Roberts PJ, *et al.* (2012) Combined PI3K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models. *Clin. Cancer Res.* 18(19):5290-5303.
- 10 18. Casado P, *et al.* (2013) Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol.* 14(4):R37.
19. Montoya A, Beltran L, Casado P, Rodriguez-Prados JC, & Cutillas PR (2011) Characterization of a TiO(2) enrichment method for label-free quantitative phosphoproteomics. *Methods* 54(4):370-378.
- 15 20. Sarbassov DD, *et al.* (2004) Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr. Biol.* 14(14):1296-1302.
21. Hay N & Sonenberg N (2004) Upstream and downstream of mTOR. *Genes Dev.* 18(16):1926-1945.
22. Kumar JK, Ping RYS, Teong HF, Goh S, & Clement M-V (2011) Activation of a non-genomic Pim-1/Bad-Pser75 module is required for an efficient pro-survival effect of Bcl-xL induced by androgen in LNCaP cells. *Int. J. Biochem. Cell. Biol.* 43(4):594-603.
- 20 23. Vincent AM & Feldman EL (2002) Control of cell survival by IGF signaling pathways. *Growth Horm. IGF Res.* 12(4):193-197.
24. Ciaccio MF, Wagner JP, Chuu C-P, Lauffenburger DA, & Jones RB (2010) Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat. Methods* 7(2):148-U195.
- 25 25. Garcia-Martinez JM, *et al.* (2009) Ku-0063794 is a specific inhibitor of the mammalian target of rapamycin (mTOR). *Biochem. J.* 421(1):29-42.
26. Munugalavadla VM, S Slaga, D Du, C Berry, L Del Rosario, G Yan, Y Boe, M Sun, L Friedman, L S Chesi, M Leif Bergsagel, P Ebens, A (2013) The PI3K inhibitor GDC-0941 combines with existing clinical regimens for superior activity in multiple myeloma. *Oncogene* 33(3):316-325.
- 30 27. Liu P, *et al.* (2011) Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nat. Med.* 17(9):1116-1120.
28. Oda K, Matsuoka Y, Funahashi A, & Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Sys. Biol.* 1:2005.0010.
- 35 29. Kanehisa M, Goto S, Sato Y, Furumichi M, & Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40(Database issue):D109-114.
30. Alcolea MP, Casado P, Rodriguez-Prados JC, Vanhaesebroeck B, & Cutillas PR (2012) Phosphoproteomic Analysis of Leukemia Cells under Basal and Drug-treated Conditions Identifies Markers of Kinase Pathway Activation and Mechanisms of Resistance. *Mol. Cell. Proteomics* 11(8):453-466.
- 40 31. Dutta B, *et al.* (2012) A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. *Br. J. Cancer* 106(6): 1107-1116.
32. Hughes TR, *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102(1):109-126
33. Muran T, *et al.* (2012) Inhibition of PI3K/mTOR leads to adaptive resistance in matrix-attached cancer cells. *Cancer Cell* 21(2):227-239.
- 45 34. Liu P, *et al.* (2011) Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nat. Med.* 17(9):1116-U1122.
35. Elkabets M, *et al.* (2013) mTORC1 Inhibition Is Required for Sensitivity to PI3K p110 alpha Inhibitors in PIK3CA-Mutant Breast Cancer. *Sci. Transl. Med.* 5(196):196ra99.
36. Klemptner SJ, Myers AP, & Cantley LC (2013) What a Tangled Web We Weave: Emerging Resistance Mechanisms to Inhibition of the Phosphoinositide 3-Kinase Pathway. *Cancer Disc.* 3(12):1345-1354.
- 50 37. Lito P, Rosen N, & Solit DB (2013) Tumor adaptation and resistance to RAF inhibitors. *Nat. Med.* 19(11): 1401-1409.
38. Gerlinger M, *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Eng. J. Med.* 366(10):883-892.
- 55 39. Le Gac M, Plucain J, Hindre T, Lenski RE, & Schneider D (2012) Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109(24):9487-9492.
40. Bolstad BM, Irizarry RA, Astrand M, & Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185-193.
41. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Molec. Biol.* 3(1):Articulo3.
- 60 42. Smyth GK (2005) *Limma: linear models for microarray data* (Springer New York).
43. R Core Team (2013) *R: A language and environment for statistical computing* (R Foundation for Statistical Computing).

**REIVINDICACIONES**

1. Un método *in vitro* para crear un conjunto de datos de sitios de modificación que se ven afectados de la misma manera por los moduladores de la misma enzima modificadora de proteínas y, por lo tanto, son indicativos de la actividad de la enzima modificadora de proteínas que está afectada por los moduladores, que comprende identificar y/o cuantificar sitios de modificación en péptidos modificados en una primera muestra que se ha tratado con un primer modulador de una enzima modificadora de proteínas y una segunda muestra que se ha tratado con un segundo modulador de la misma enzima modificadora de proteínas usando espectrometría de masas (EM); y agrupar sitios de modificación en péptidos modificados de la primera muestra que se ha tratado con el primer modulador de una enzima modificadora de proteínas y sitios de modificación en péptidos modificados de la segunda muestra que se ha tratado con el segundo modulador de la misma enzima modificadora de proteínas en un solo grupo, de acuerdo con el efecto de dichos primer y segundo moduladores de dicha enzima modificadora de proteínas sobre dichos sitios de modificación, en donde los sitios de modificación en péptidos modificados de una primera muestra y los sitios de modificación en péptidos modificados de una segunda muestra se colocan en un solo grupo basado en los sitios de modificación que se ven afectados de la misma manera por los moduladores primero y segundo de la enzima modificadora de proteínas, y en donde dichos primer y segundo moduladores de dicha enzima modificadora de proteínas son diferentes.
2. El método de acuerdo con la reivindicación 1, en donde dichos sitios de modificación se seleccionan del grupo que consiste en sitios de fosforilación, sitios de acetilación, sitios de glicosilación, sitios de metilación y sitios de lipidación.
3. El método de acuerdo con la reivindicación 1 o 2, en donde dicha primera muestra y dicha segunda muestra son de:
- (a) un fluido corporal, opcionalmente orina o sangre; o
  - (b) un tejido, opcionalmente un tejido canceroso, opcionalmente un tumor.
4. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde dicha enzima modificadora de proteínas se selecciona del grupo que consiste en una proteína quinasa, proteína fosfatasa, proteína glicosiltransferasa, proteína acetiltransferasa, proteína metiltransferasa y proteína palmitoiltransferasa.
5. El método de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde dicho modulador de una enzima modificadora de proteínas es
- (a) una molécula pequeña, ARNi, péptido terapéutico o anticuerpo; y/o
  - (b) un inhibidor de una enzima modificadora de proteínas.
6. El método de acuerdo con la reivindicación 5, en donde dicho inhibidor de una enzima modificadora de proteínas es un inhibidor de quinasas.
7. El método de acuerdo con la reivindicación 6, en donde dicho inhibidor de quinasas es un inhibidor de una quinasa seleccionada del grupo que consiste en Akt, CAMK2, EGFR, ERK, MEK, mTOR, p70S6K, PI3K, PKC y ROCK.
8. El método de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde dicho efecto de dichos primer y segundo moduladores de dicha enzima modificadora de proteínas en dicho sitio de modificación es una reducción o aumento en la abundancia de dicho sitio de modificación en comparación con una muestra de control.
9. Un método para preparar una base de datos, que comprende crear un conjunto de datos de sitios de modificación de acuerdo con un método de una cualquiera de las reivindicaciones precedentes y recopilar dicho conjunto de datos en una base de datos.
10. El método de acuerdo con la reivindicación 9, en donde dicha base de datos comprende además información sobre las identidades de proteínas que contienen los sitios de modificación, el tipo de modificación, el tipo de muestra en la que está presente el sitio de modificación y/o el modulador que aumenta o disminuye la modificación en el sitio de modificación.
11. Un método *in vitro* para diagnosticar una enfermedad que comprende:
- (a) Crear un conjunto de datos de prueba de sitios de modificación mediante:
    - i. El tratamiento de una o más muestras de prueba de un sujeto con uno o más moduladores de una enzima modificadora de proteínas,
    - ii. La identificación y/o la cuantificación de sitios de modificación en péptidos modificados en una o más muestras de prueba,
    - iii. El agrupamiento de sitios de modificación en péptidos modificados de una o más muestras de prueba en un solo grupo, creando así el conjunto de datos de prueba de sitios de modificación, y

iv. opcionalmente la recopilación del conjunto de datos de prueba en una base de datos de prueba,

(b) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 8 o comparar la base de datos de prueba con una base de datos preparada de acuerdo con las reivindicaciones 9 o 10,

(c) Encontrar una diferencia significativa entre el conjunto de datos de prueba o la base de datos de prueba y el conjunto de datos de control o la base de datos de control,

(d) Diagnosticar que el sujeto tiene la enfermedad si se encuentra una diferencia significativa entre el conjunto de datos o la base de datos en la etapa (c).

12. El método de la reivindicación 11, en donde la enfermedad es cáncer.

13. Un método para identificar uno o más biomarcadores para la actividad de una enzima modificadora de proteínas que comprende crear un conjunto de datos de sitios de modificación de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 8 o preparar una base de datos de acuerdo con las reivindicaciones 9 o 10, en donde los sitios de modificación en el único grupo se identifican como el uno o más biomarcadores para la actividad de una enzima modificadora de proteínas.

14. Un método *in vitro* para determinar que una sustancia de prueba es un modulador de una enzima modificadora de proteínas que comprende tratar una muestra de prueba con la sustancia de prueba, identificar y/o cuantificar sitios de modificación en péptidos modificados en la muestra de prueba, crear un conjunto de datos de prueba de sitios de modificación a partir de los sitios de modificación identificados y/o cuantificados en la muestra de prueba, que comprende además:

(a) Comparar el conjunto de datos de prueba con un conjunto de datos de sitios de modificación creado de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 8 y/o

(b) Recopilar dicho conjunto de datos de prueba en una base de datos de prueba y comparar la base de datos de prueba con una base de datos preparada de acuerdo con las reivindicaciones 9 o 10.

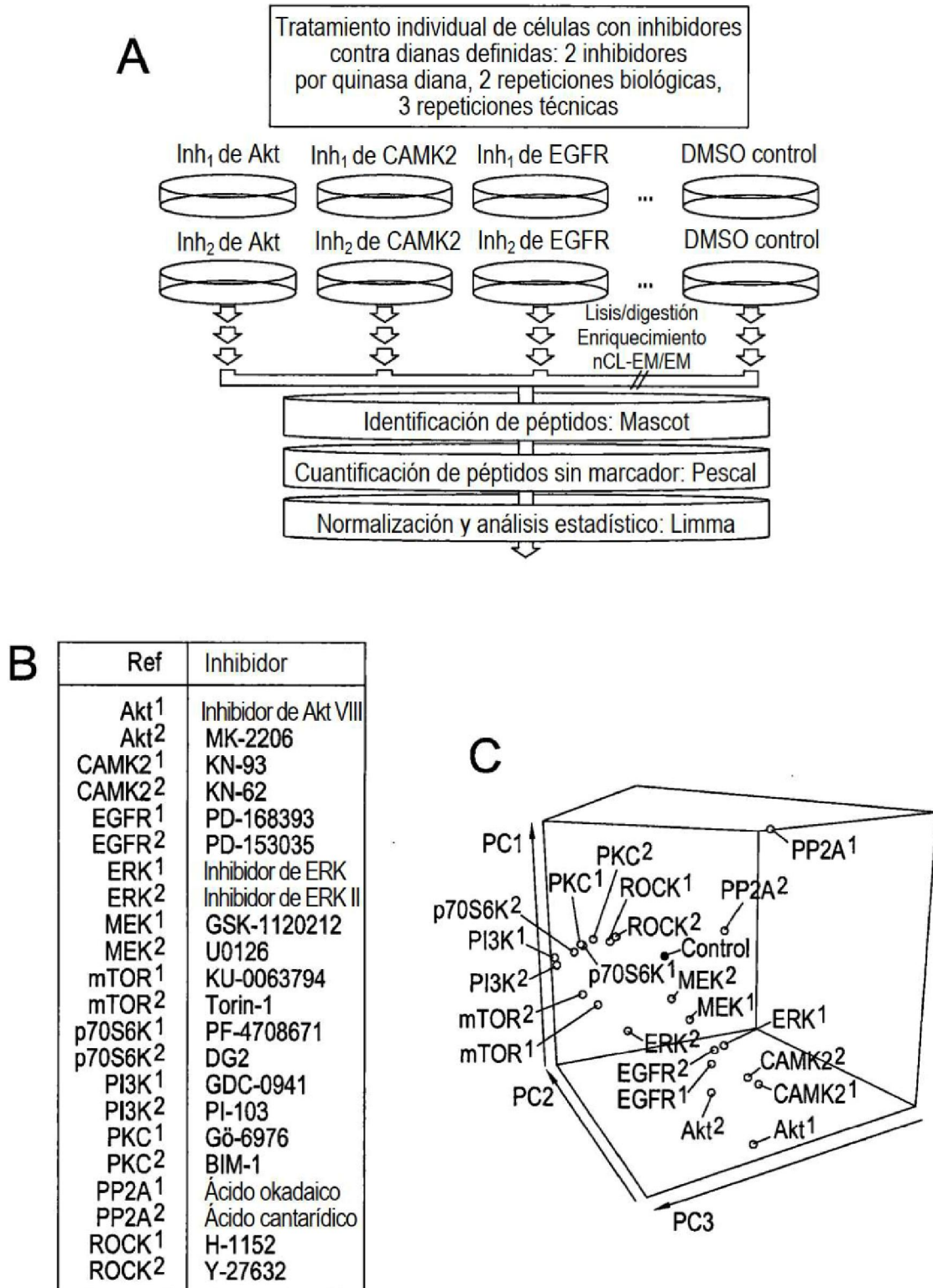
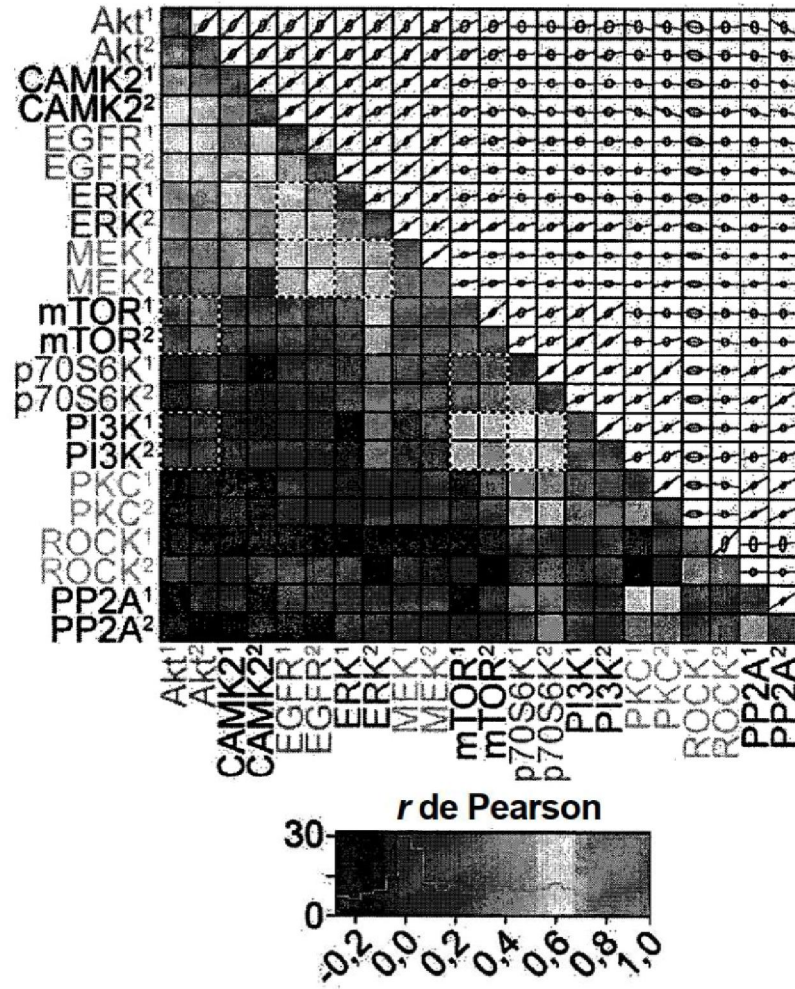


Fig. 1

D



E



Fig. 1

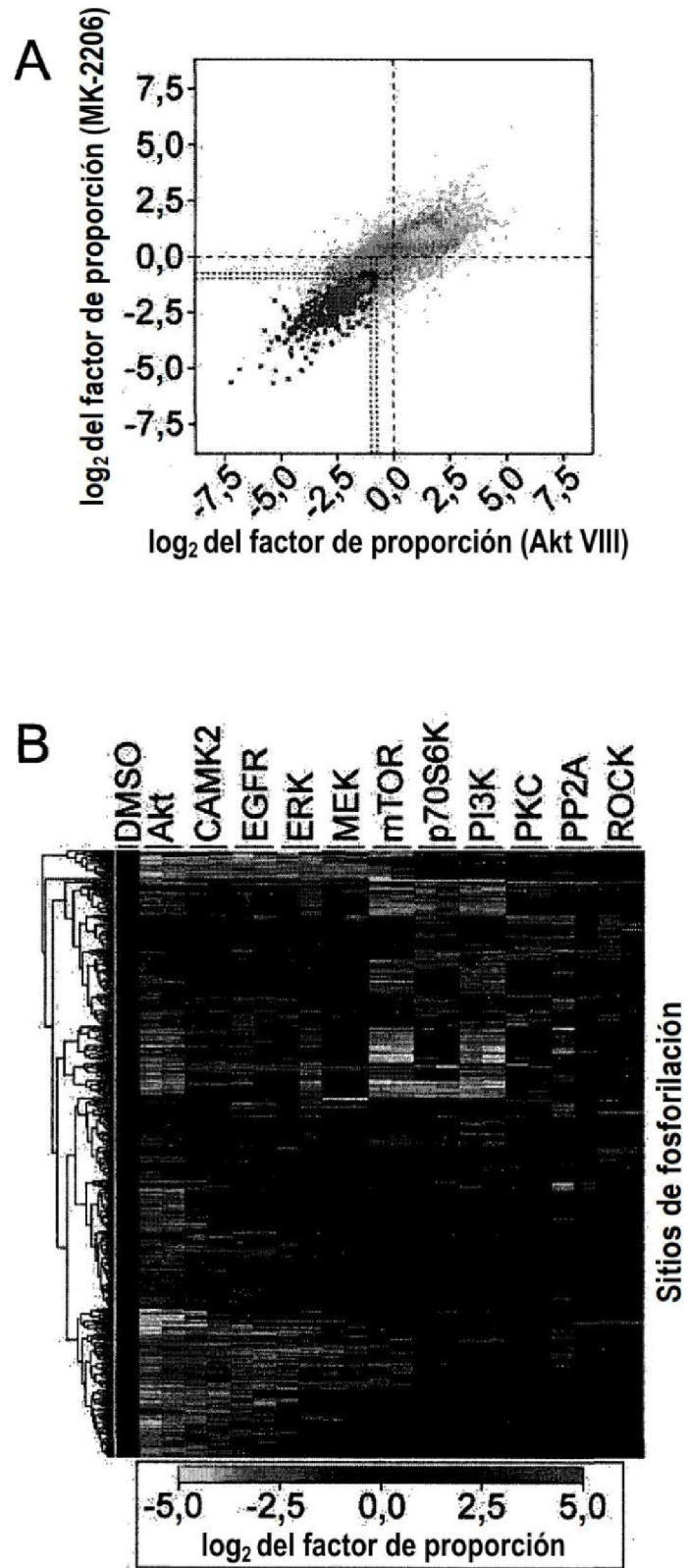


Fig. 2

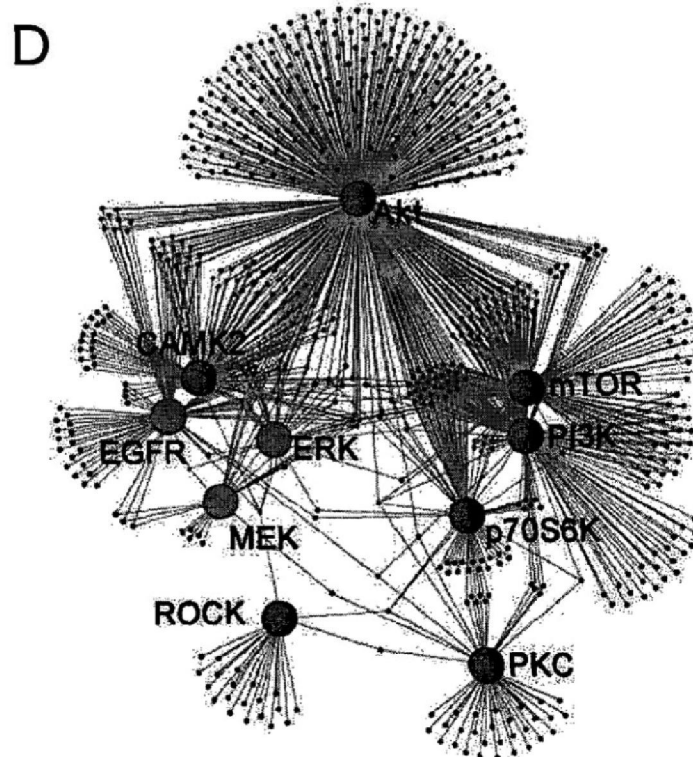
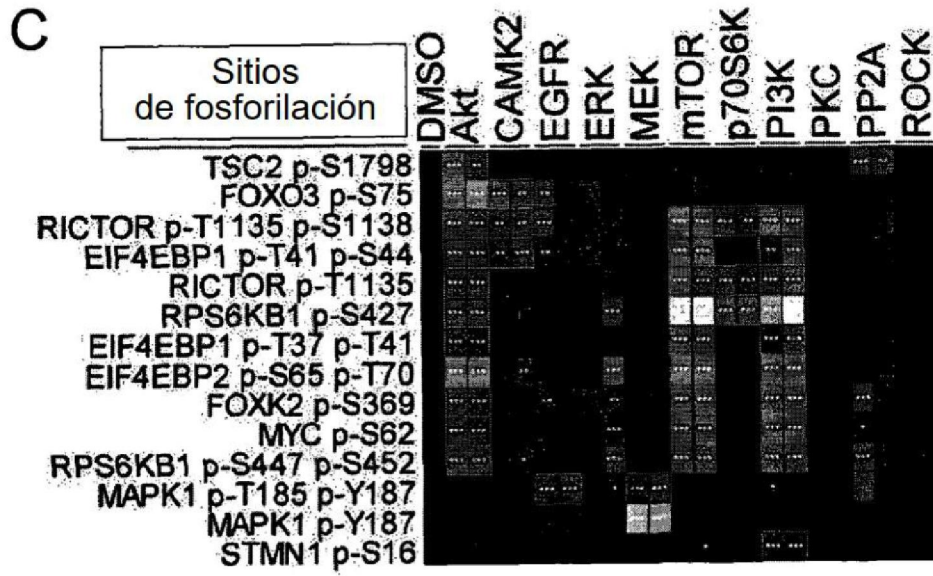


Fig. 2

A

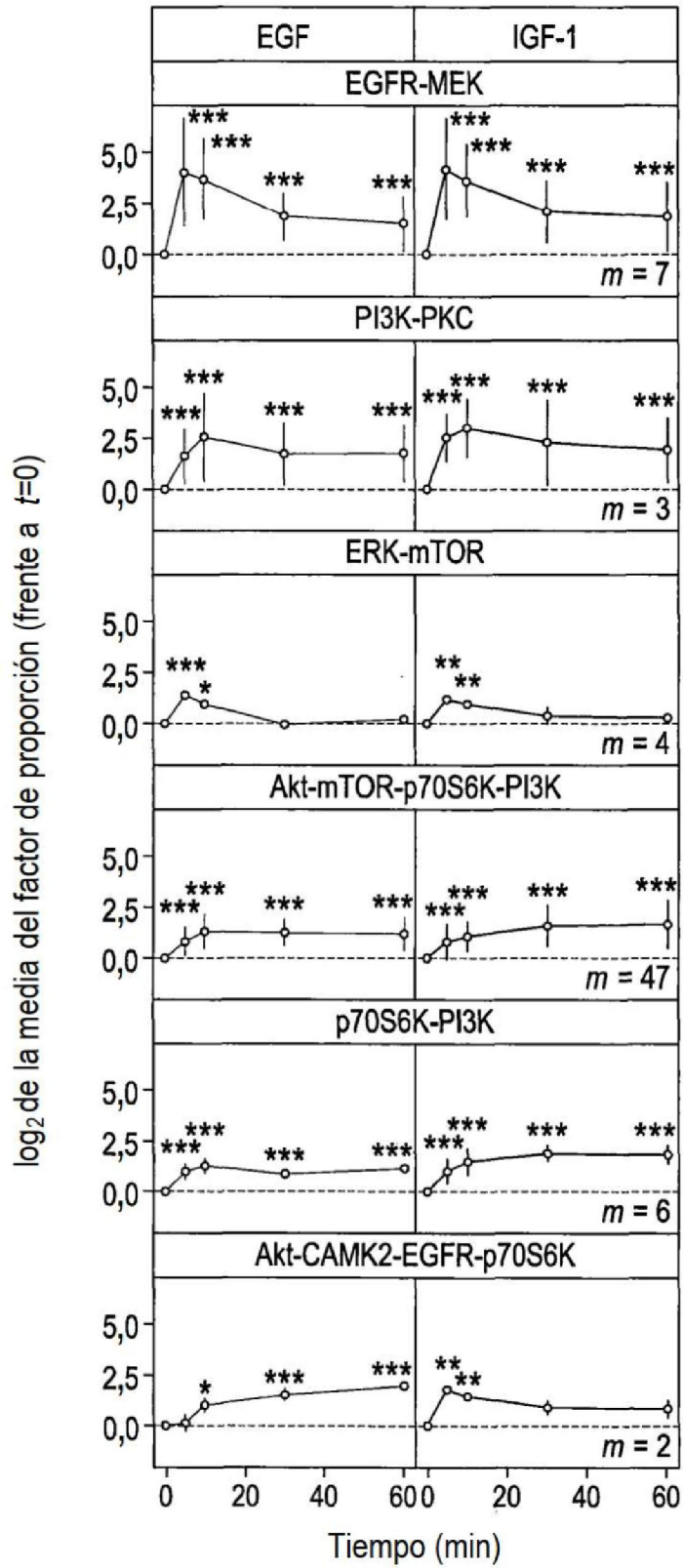
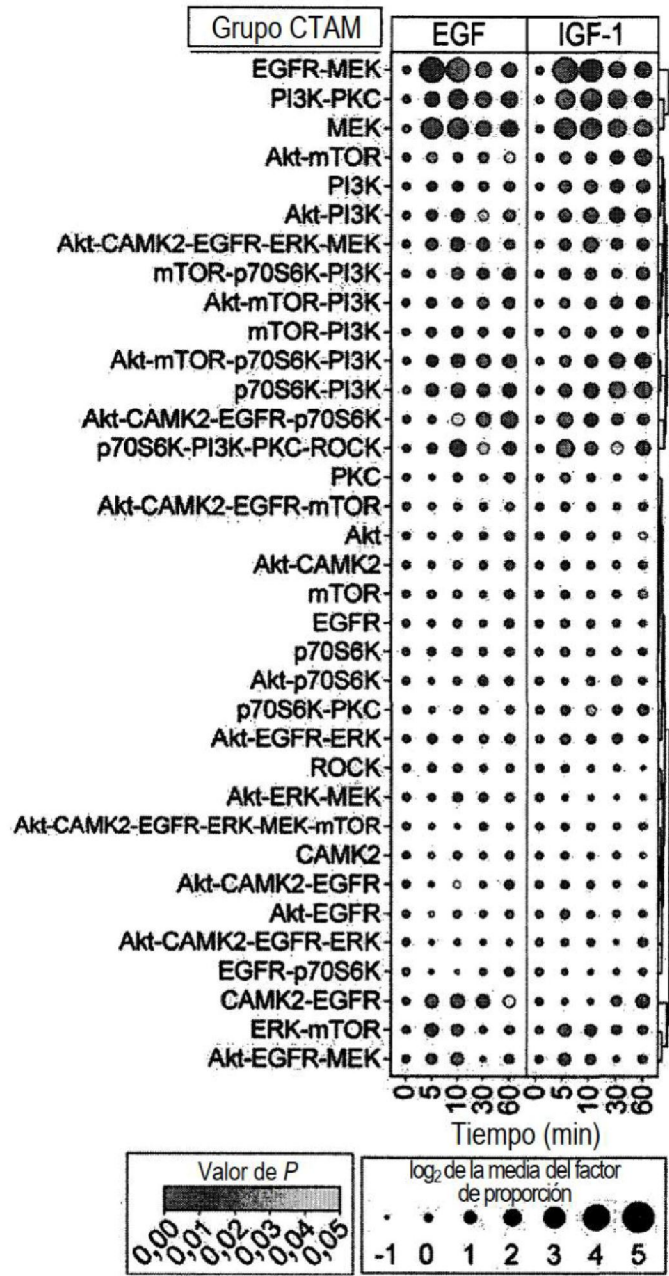


Fig. 3

**B**



**C**

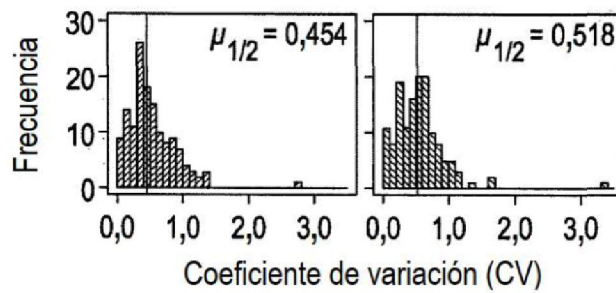
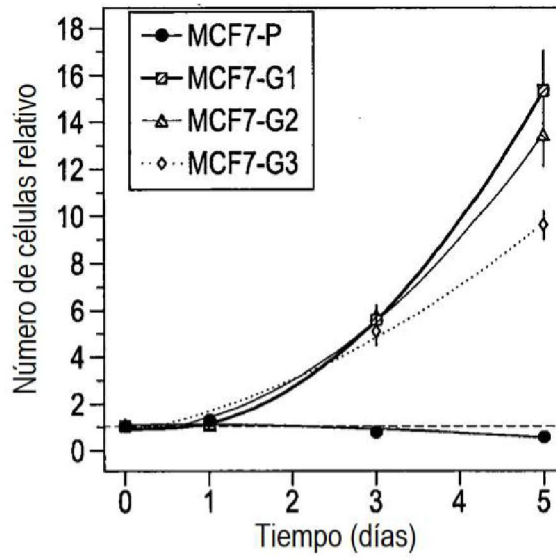
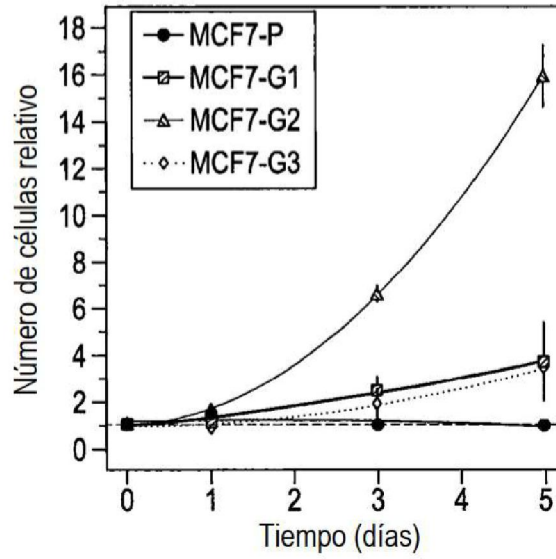


Fig. 3

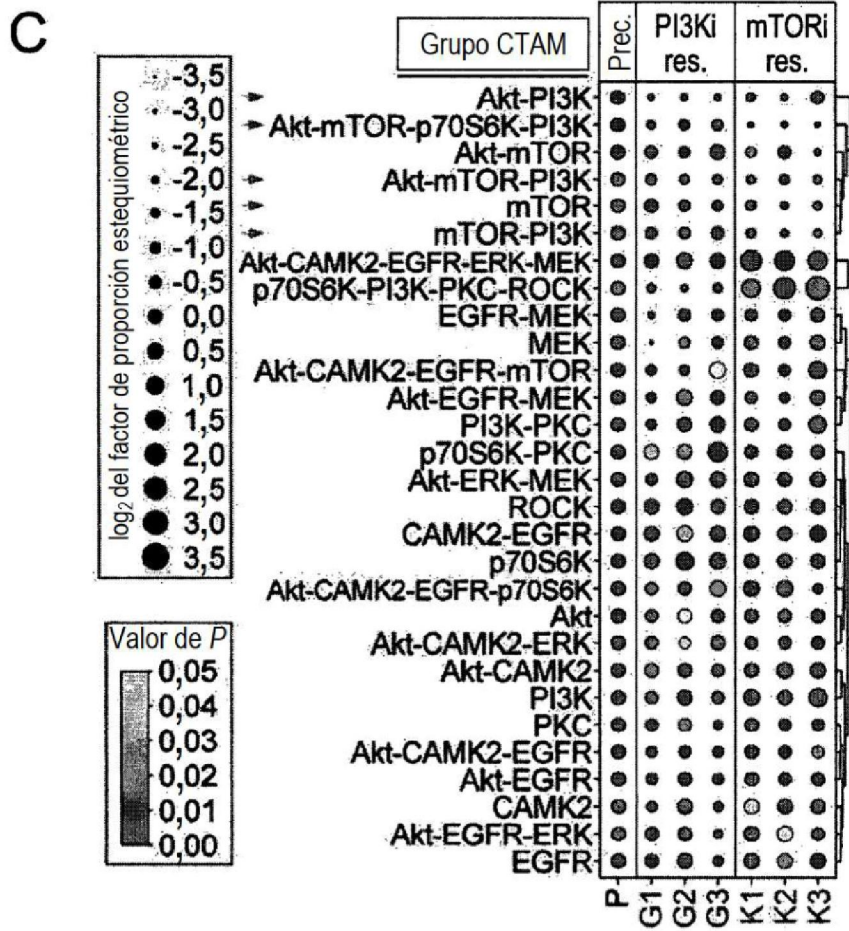
**A**



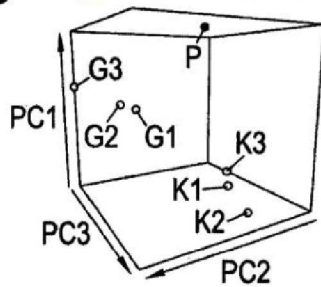
**B**



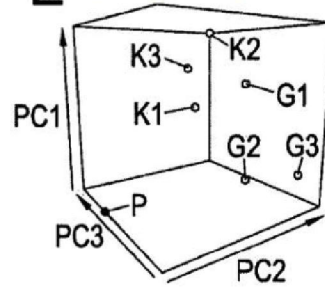
**Fig. 4**



**D** Entrada: datos de EM



**E** Entrada: datos de MTS



Entrada: datos con CV

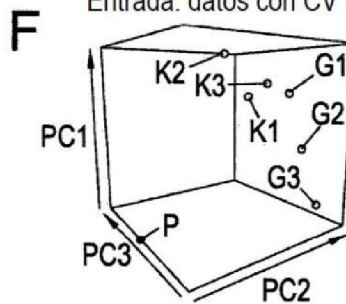


Fig. 4

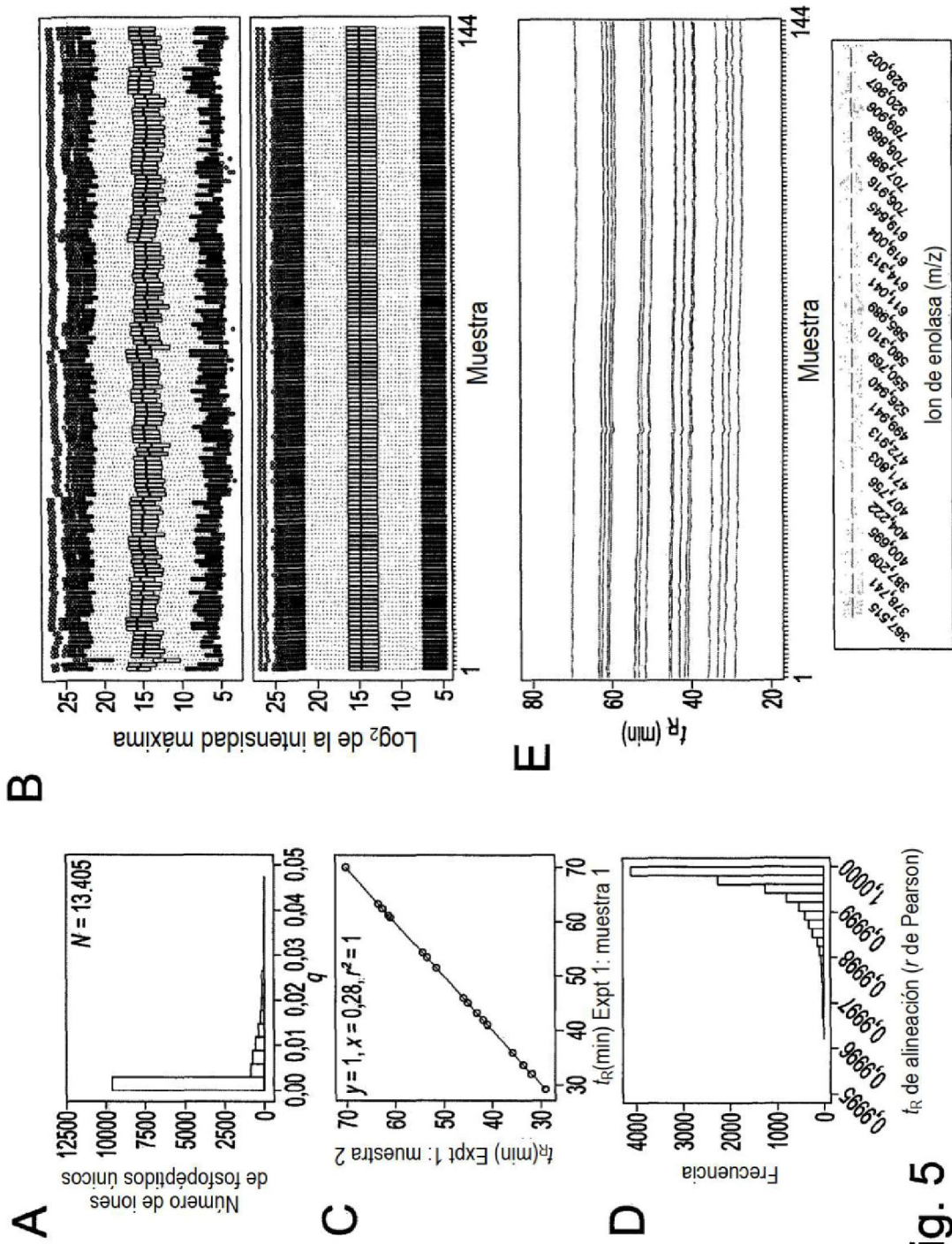


Fig. 5

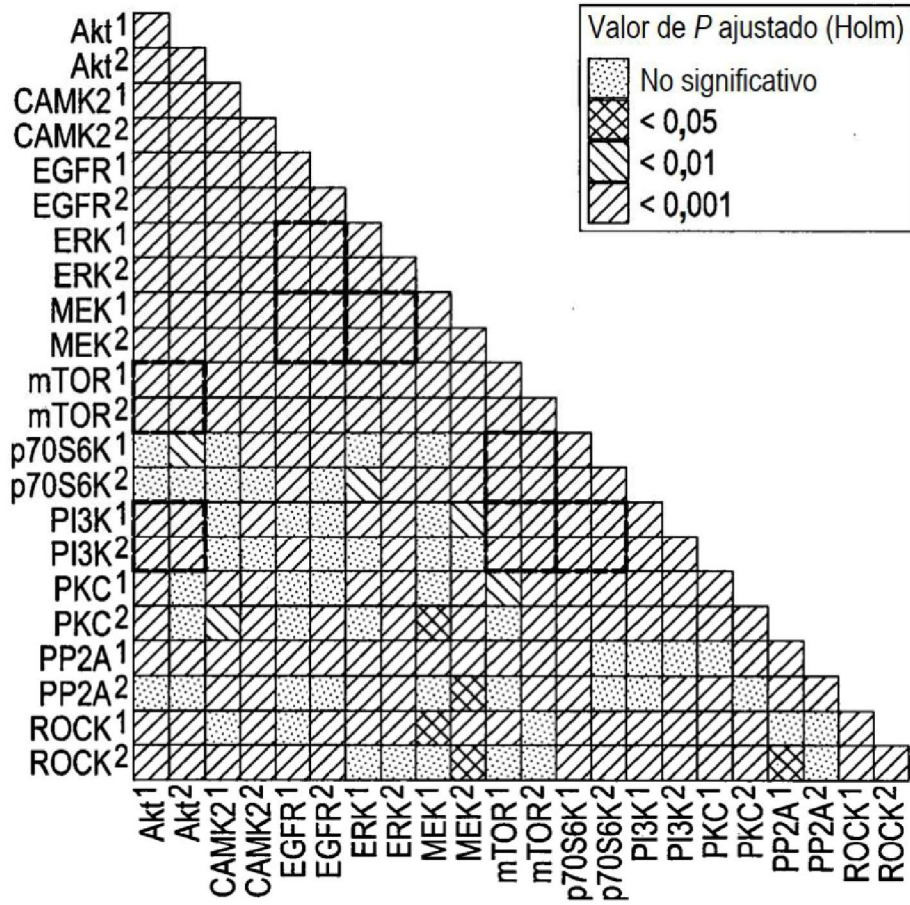


Fig. 6

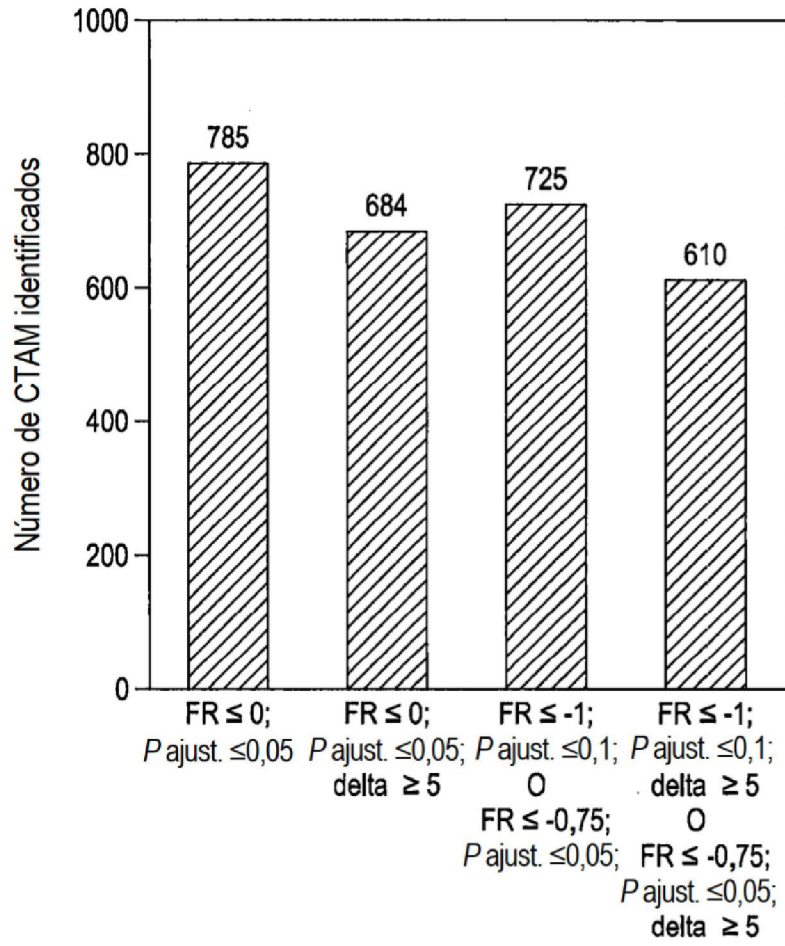
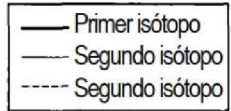


Fig. 7

Sustrato de AKT1 rico en Prolina 1 (p-S183),  
 [S]LPVSVPVWGFK, AKT1S1 183 - 194, Fosfo (ST),  
 Puntuación Mascot = 70,1, puntuación delta = 40,77,  $m/z = 698,3588$ ,  $z = 2$ ,  $t_R \sim 80$  min



Repetición biológica 1; repeticiones técnicas a-c      Repetición biológica 2; repeticiones técnicas a-c

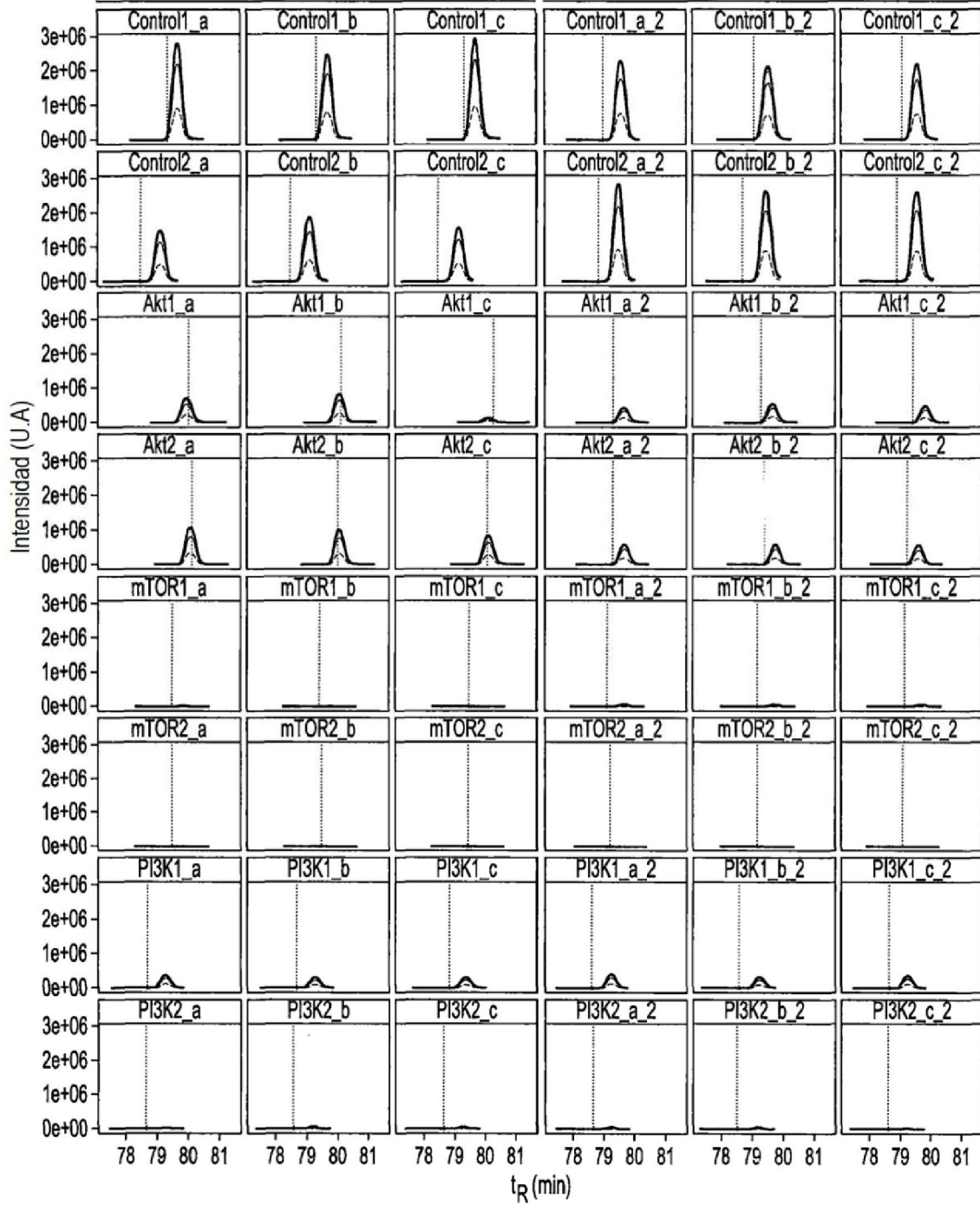


Fig. 8

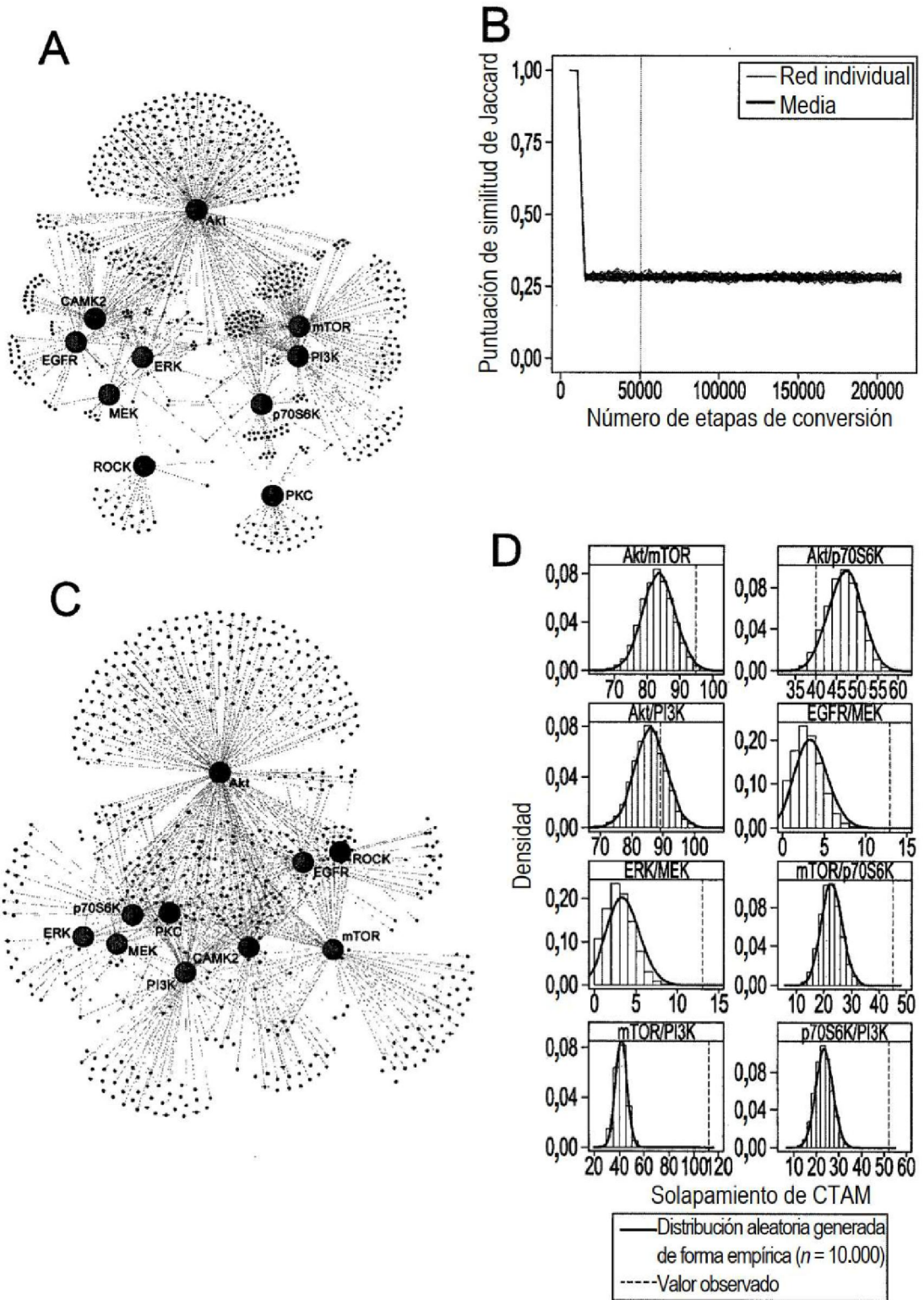
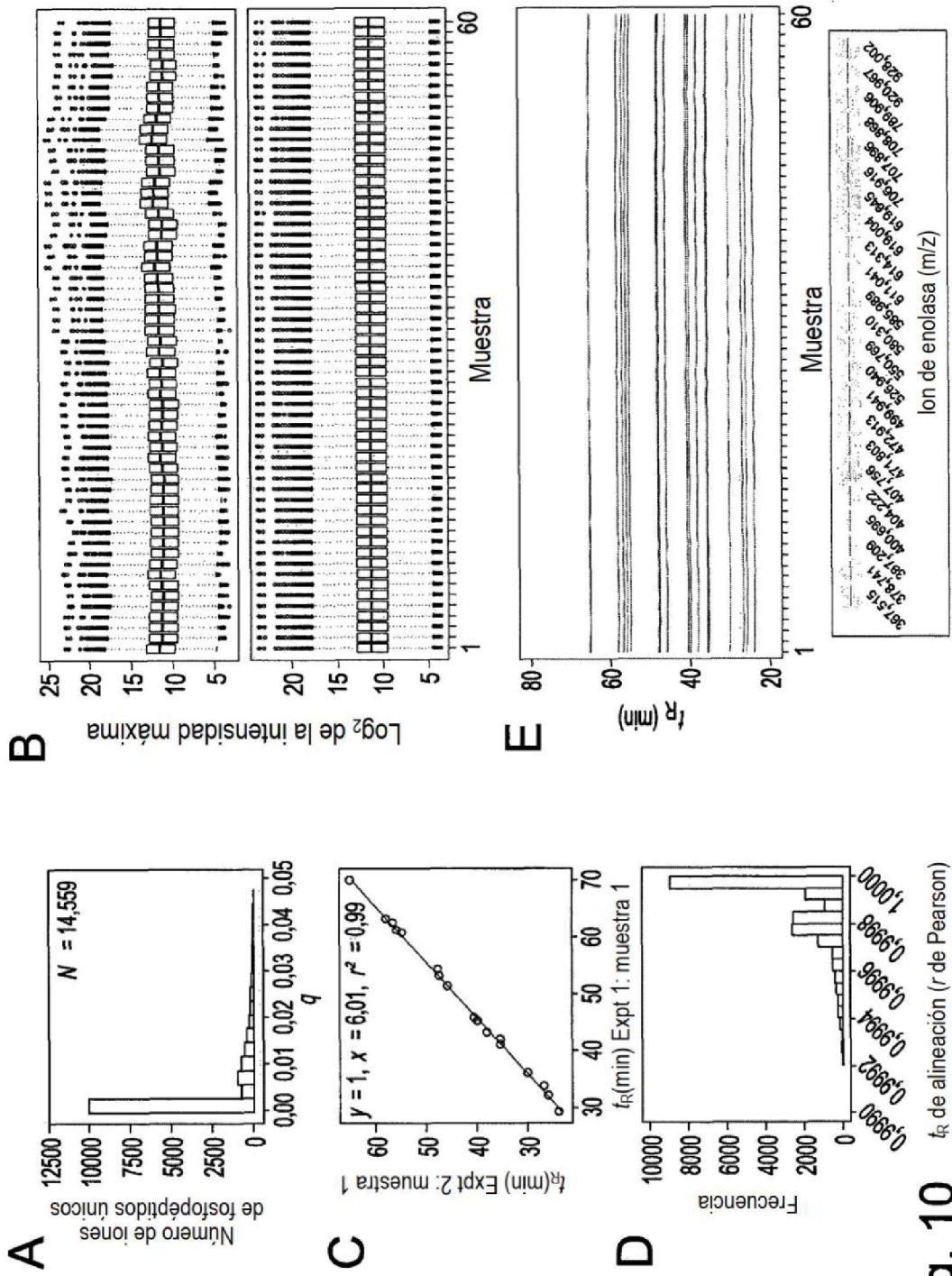


Fig. 9



**Fig. 10**  $t_R$  de alineación ( $r$  de Pearson)

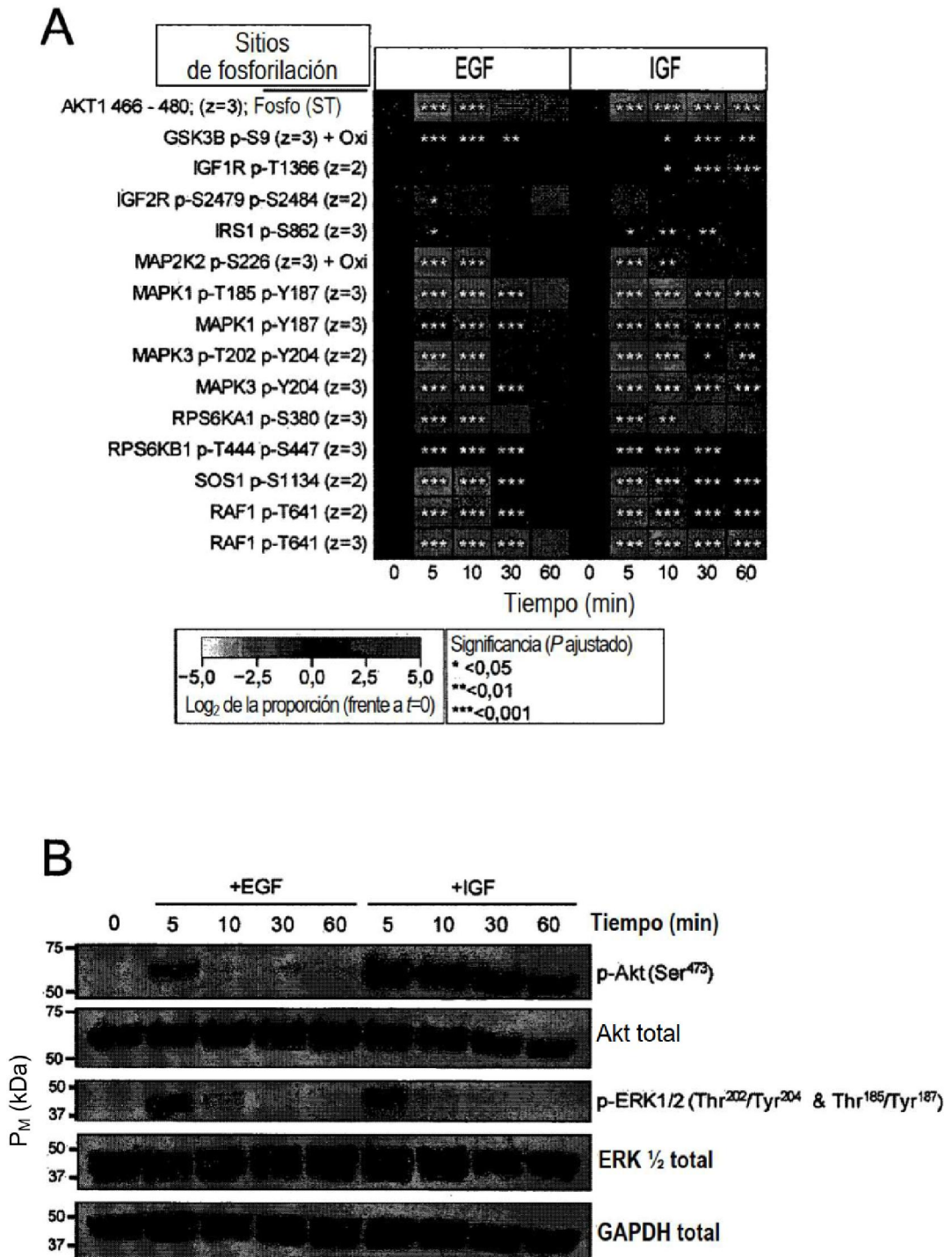
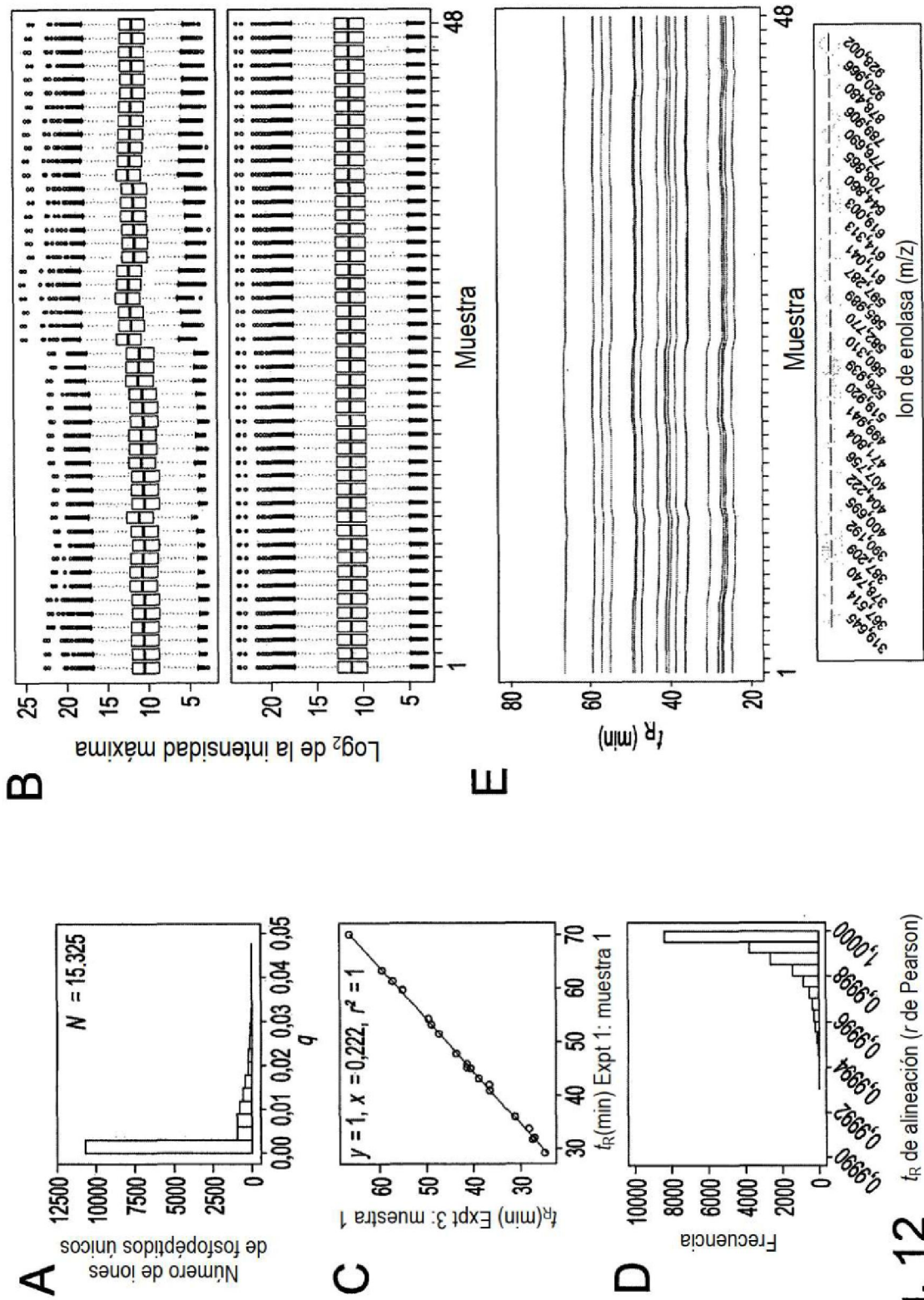


Fig. 11



**Fig. 12**  $t_R$  de alineación ( $r$  de Pearson)

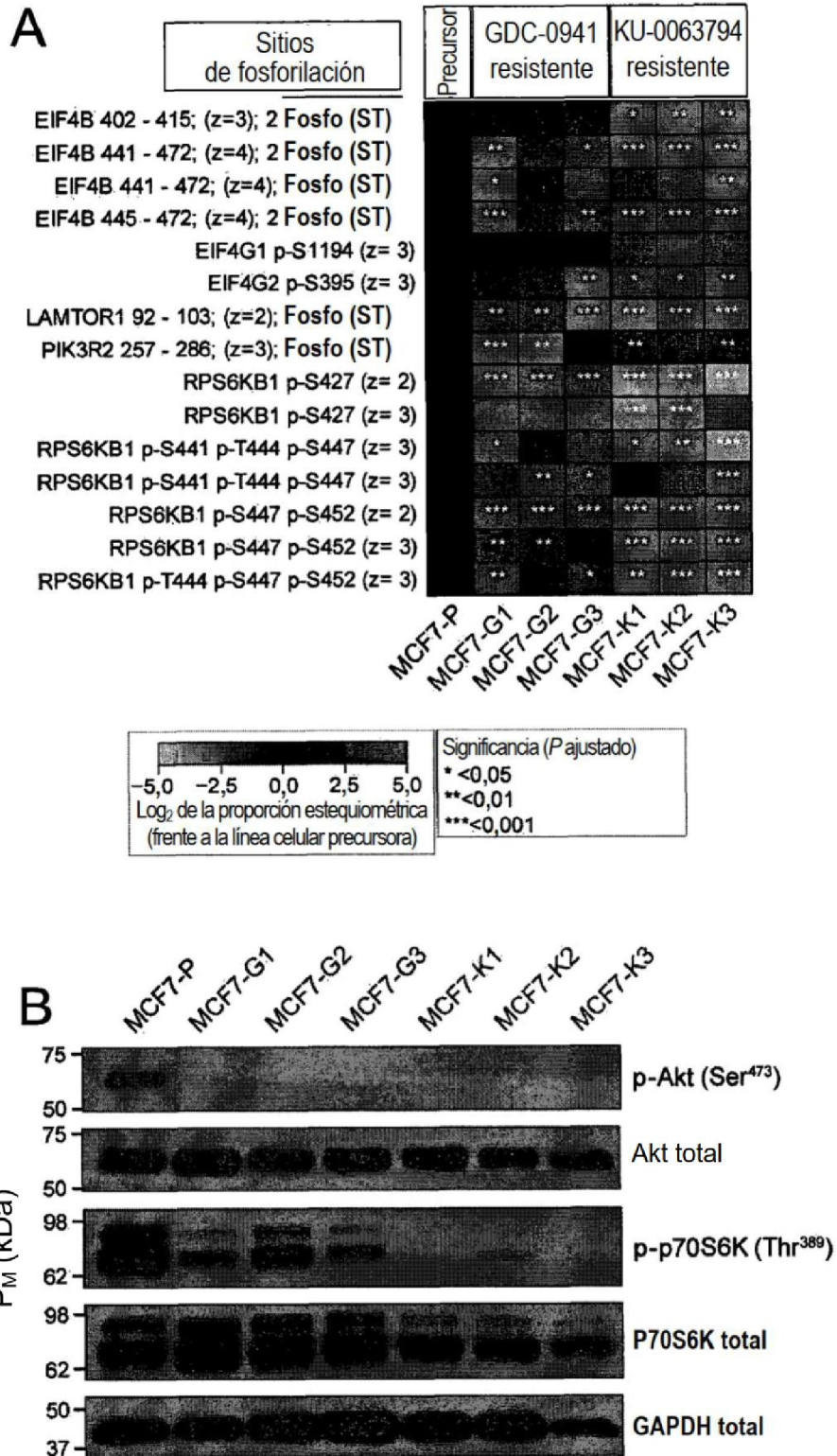


Fig. 13

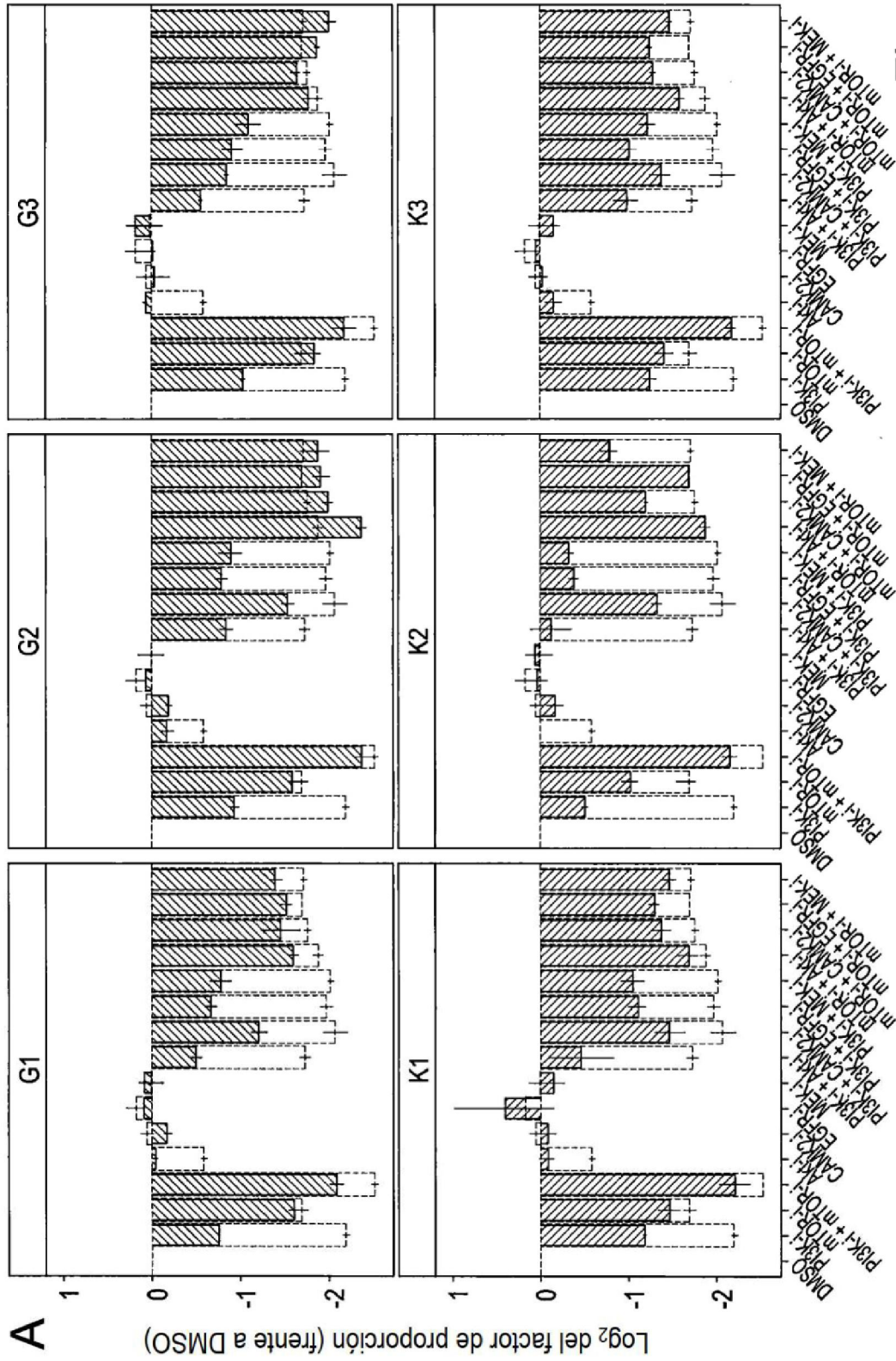


Fig. 14

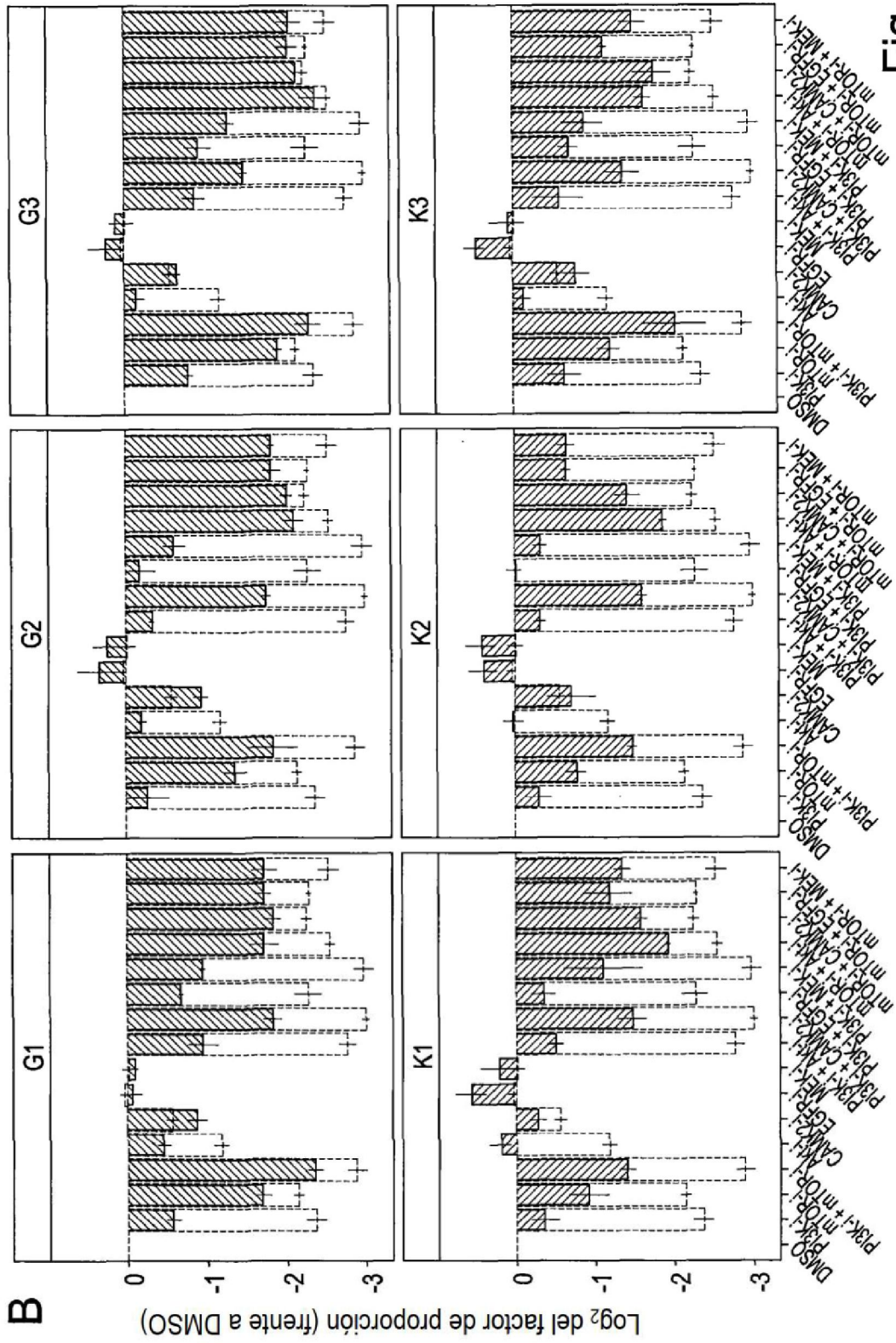


Fig. 14